

GenCore version 5.1.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 10, 2003, 11:44:43 ; Search time 2646.23 Seconds
(without alignments)
11085.832 Million cell updates/sec

Title: US-09-923-236-1
Perfect score: 1008
Sequence: 1 agacagactaaaaagccat.....tagataatttagaccaatgg 1008

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl.*

- 1: gb_ba.*
- 2: gb_htg.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vi.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_om.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pl.*
- 26: em_ro.*
- 27: em_sts.*
- 28: em_un.*
- 29: em_vi.*
- 30: em_htg_hum.*
- 31: em_htg_inv.*
- 32: em_htg_other.*
- 33: em_htg_mus.*
- 34: em_htg_pln.*
- 35: em_htg_rod.*
- 36: em_htg_man.*
- 37: em_htg_vrt.*
- 38: em_sy.*
- 39: em_htgo_hum.*
- 40: em_htgo_mus.*
- 41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	997	98.9	998	6	AX061225	Sequence
2	578.4	57.4	96481	9	AC104811	Homo sapi
3	575.2	57.1	157836	2	AC060228	Homo sapi
c	537.6	53.3	185969	2	AC069037	Homo sapi
5	419.4	41.6	185969	2	AC069037	Homo sapi
c	122.6	12.2	177035	2	AC099443	Rattus no
7	121.2	12.0	3489	6	AR072952	Sequence
c	121.2	12.0	3489	6	AX107938	Sequence
c	121.2	12.0	3489	14	KS052064	U52064 Kaposi's sa
10	121.2	12.0	32207	6	AR065852	Sequence
11	121.2	12.0	32207	6	AR127850	Sequence
12	121.2	12.0	32207	6	AR194752	Sequence
13	121.2	12.0	137508	14	KS075698	Kaposi's sa
14	117.8	11.7	28559	14	AF148805	Kaposi's sa
15	115.2	11.4	133661	14	U93872	Kaposi's sa
16	111.8	11.1	225045	2	AL833803	Mus muscu
c	110.8	11.0	2931	14	AF192756	Kaposi's
18	110.6	11.0	5719	10	MMCATS1	Mus muscu
c	110.6	11.0	18876	10	AC092203	Mus muscu
c	110.4	11.0	2286	6	AX460942	Sequence
21	110.4	11.0	168210	2	AC116700	Mus muscu
22	109.6	10.9	228404	2	AC095672	Rattus no
23	109.4	10.9	272545	2	AC090533	Mus muscu
c	109	10.8	182534	10	AC074046	Mus muscu
25	108.4	10.8	155662	2	AC112388	Rattus no
26	108	10.7	186559	10	AL608742	Mouse DNA
27	107.8	10.7	184865	2	AL807824	Mus muscu
28	107.8	10.7	227588	2	AC094571	Rattus no
c	107.2	10.6	176030	2	AC129650	Rattus no
30	107.2	10.6	201470	2	AC113649	Rattus no
c	107.2	10.6	287058	2	AC095595	Rattus no
32	106.6	10.6	190444	2	AC078945	Mus muscu
33	106.6	10.6	225004	2	AC099725	Mus muscu
c	106	10.5	5120	6	AX067148	Sequence
c	106	10.5	5120	9	AF152102	Homo sapi
36	105.8	10.5	287323	2	AC093363	Mus muscu
37	105.6	10.5	153068	2	AC119703	Rattus no
38	105.6	10.5	180418	2	AC106674	Rattus no
c	105.6	10.5	185991	2	AC130124	Rattus no
c	104.8	10.4	2479	10	RN087960	Rattus norv
c	104.8	10.4	3941	10	RNLAR2	X83546 R.norvegic
42	104.8	10.4	190289	14	MCU60315	Molluscum c
c	104.4	10.4	3012	14	AF360120	Human her
44	104.2	10.3	1442	9	AF087653	Homo sapi
45	104.2	10.3	204259	2	AC110817	Mus muscu

ALIGNMENTS

RESULT 1
AX061225
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS

AX061225
Sequence
AX061225
AX061225.1
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 998)
Lal,P., Yang,J., Yue,H., Hillman,J.L., Tang,Y.T., Bandman,O.,
Burford,N., Baughn,M.R., Azimzai,Y., Lu,D.A., Au-Young,J. and
Patterson,C.

AX061225
Sequence
AX061225
AX061225.1
GI:12406361
72 from Patent WO0078953.
linear
PAT 22-JAN-2001

TITLE

Human transport proteins

JOURNAL

Patent: WO 0078953-A 72 28-DEC-2000;

Incyte Genomics, Inc. (US)

Location/Qualifiers

1..998

source

/organism="Homo sapiens"

/db_xref="taxon:9606"

SE COUNT

253 a 249 c 190 g 306 t

IGIN

Query Match

98.9%;

Score 997;

DB 6;

Length 998;

Best Local Similarity

100.0%;

Pred. No. 9.9e-193;

Matches 997;

Conservative 0;

Mismatches 0;

Indels 0;

Gaps 0;

6

GACTAAAGCCATGATTTCTTCGTTCTCTCTAAAGAGAAAAATATAATTTAAAA

65

1

GACTAAAGCCATGATTTCTTCGTTCTCTCTAAAGAGAAAAATATAATTTAAAA

60

66

ATACATTGCGTATTTCTTAAACATAAATTTATAGTGTTAATATTTCATAGGGTCAATCA

125

61

ATACATTGCGTATTTCTTAAACATAAATTTATAGTGTTAATATTTCATAGGGTCAATCA

120

126

AAATGAAGCTTCTCCTTTGGCGCTGCATGTATGTTGCTTTTGCAGGAAGACAGCGT

185

121

AAATGAAGCTTCTCCTTTGGCGCTGCATGTATGTTGCTTTTGCAGGAAGACAGCGT

180

186

TCCCCTTCATTGGTGAGATGACATGACGATGGTGCACCCACTTCATCTCTCGAATA

245

181

TCCCCTTCATTGGTGAGATGACATGACGATGGTGCACCCACTTCATCTCTCGAATA

240

246

TTCTTATGGCATACGGAATTTACACCTCTCTTTATATCGCCAGTGAATACAGTCC

305

241

TTCTTATGGCATACGGAATTTACACCTCTCTTTATATCGCCAGTGAATACAGTCC

300

306

CCAGTTACCTGGGAATFACTTACACTGACACAGGTTACCTTCGTATCCCTGGATTCTAA

365

301

CCAGTTACCTGGGAATFACTTACACTGACACAGGTTACCTTCGTATCCCTGGATTCTAA

360

366

CTTCTCCTGGATTCCCTCTATGTATACATCCGTTGGTTTCCCTTAGCTACTCAGTTGA

425

361

CTTCTCCTGGATTCCCTCTATGTATACATCCGTTGGTTTCCCTTAGCTACTCAGTTGA

420

426

ATGTTCTCCTCTCCTCCTTAGGGGTTTCCGTTTGTGCCCTCCTTCAAGGTTTTTTTCAG

485

421

ATGTTCTCCTCTCCTCCTTAGGGGTTTCCGTTTGTGCCCTCCTTCAAGGTTTTTTTCAG

480

486

CAGCTGCAGCACCCCTGCCCCACCTATTGACAGCTGAGCCTGCTGCAGCTGCACCTCTTA

545

481

CAGCTGCAGCACCCCTGCCCCACCTATTGACAGCTGAGCCTGCTGCAGCTGCACCTCTTA

540

546

CAGCCACACCTGTAGCAGCTGAGCTGTGTCAGGGGCCCTGTGTCAGCTGAGCCTGCTG

605

541

CAGCCACACCTGTAGCAGCTGAGCTGTGTCAGGGGCCCTGTGTCAGCTGAGCCTGCTG

600

606

CAGAGGCACCTGTTGGAGCTGAGCCTGTGTCAGAGGCACCTGTTGTCAGCTGAGCCTGCTG

665

601

CAGAGGCACCTGTTGGAGCTGAGCCTGTGTCAGAGGCACCTGTTGTCAGCTGAGCCTGCTG

660

666

CAGAGGCACCTGTTGGAGTGGAGCCAGCTGCAGAGGACCTTTCACAGCTGAGCCTGCTA

725

661

CAGAGGCACCTGTTGGAGTGGAGCCAGCTGCAGAGGACCTTTCACAGCTGAGCCTGCTA

720

726

CAGCCAAAGCTGTGCCGCCAGAACTCACCCTTCTCCCTCTCTTTGAACAGGCAATCAGT

785

721

CAGCCAAAGCTGTGCCGCCAGAACTCACCCTTCTCCCTCTCTTTGAACAGGCAATCAGT

780

786

GAATTTCTCTAGAAGTACCATTGGTTCAATTCTATCTACTGATGCAGAAATAAGTGAAT

845

781

GAATTTCTCTAGAAGTACCATTGGTTCAATTCTATCTACTGATGCAGAAATAAGTGAAT

840

846

CTACAAAAGTTTTCTTTCTTTTCCAAAGACTATTTTCATTCGTGTTATTCAGAGTATTC

905

841

CTACAAAAGTTTTCTTTCTTTTCCAAAGACTATTTTCATTCGTGTTATTCAGAGTATTC

900

Qy

906

TCTCACTACATTGATTTGTTGGTAGTTTTCCTTGGACTTAATTTATATTTGAAAAA

965

Db

901

TCTCACTACATTGATTTGTTGGTAGTTTTCCTTGGACTTAATTTATATTTGAAAAA

960

Qy

966

CATTGATAATTAATAAATAAATAGATAATTTAGAC

1002

Db

961

CATTGATAATTAATAAATAAATAGATAATTTAGAC

997

RESULT 2

AC104811

LOCUS

AC104811

96481 bp

DNA

linear

PRI 16-APR-2002

DEFINITION

Homo sapiens BAC clone RP11-653L5 from 4, complete sequence.

ACCESSION

AC104811

VERSION

AC104811.4

GI:19698738

HTG.

KEYWORDS

HTG.

SOURCE

Homo sapiens.

ORGANISM

Homo sapiens

REFERENCE

1 (bases 1 to 96481)

AUTHORS

Sulston,J.E. and Waterston,R.

TITLE

Toward a complete human genome sequence

JOURNAL

Genome Res. 8 (11), 1097-1108 (1998)

MEDLINE

99063792

PUBMED

9847074

REFERENCE

2 (bases 1 to 96481)

AUTHORS

Desai,A., Haakenson,W. and Dignan,G.

TITLE

The sequence of Homo sapiens BAC clone RP11-653L5

JOURNAL

Unpublished (2001)

REFERENCE

3 (bases 1 to 96481)

AUTHORS

Waterston,R.H.

TITLE

Direct Submission

JOURNAL

Submitted (21-DEC-2001) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA

REFERENCE

4 (bases 1 to 96481)

AUTHORS

Waterston,R.H.

TITLE

Direct Submission

JOURNAL

Submitted (14-MAR-2002) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA

REFERENCE

5 (bases 1 to 96481)

AUTHORS

Waterston,R.H.

TITLE

Direct Submission

JOURNAL

Submitted (24-MAR-2002) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA

REFERENCE

6 (bases 1 to 96481)

AUTHORS

Waterston,R.H.

TITLE

Direct Submission

JOURNAL

Submitted (29-MAR-2002) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA

REFERENCE

7 (bases 1 to 96481)

AUTHORS

Waterston,R.

TITLE

Direct Submission

JOURNAL

Submitted (16-APR-2002) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

COMMENT

On Mar 24, 2002 this sequence version replaced gi:19424658.

Center: Washington University Genome Sequencing Center

Center code: WUGSC

Web site: <http://genome.wustl.edu/gsc>

Contact: sapiens@wustl.wustl.edu

----- Summary Statistics -----

Center project name: H_NH0653L05

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap

Db	60879	TTAGGATGACAATGACGATGGTCACCCACCTTCATCCATCTCTGAATATTCTCTTATGGCAT	60938
Qy	259	ACGGAAATTTACCACCTCCTCTTTATTATATCGCCACAGTGAATACAGTCCCCAGTTACCCCTGG	318
Db	60939	ACGGAAATTTACCACCTCCTCTTTATTATATCGCCACAGTGAATACAGTCCCCAGTTACCCCTGG	60998
Qy	319	GAATACACTACACTCACAGAGTTTACCTTCGTATCCCTCGGATTCTAACTTCTCCTTGGATT	378
Db	60999	GAATACACTACACTCACAGAGTTTACCTTCGTATCCCTCGGATTCTAACTTCTCCTTGGATT	61058
Qy	379	CCCTATGCTATCACATCCCGTGGTTTTCCCTTAGCTACTCAGTTGAATGTTCTCCTCCTCT	438
Db	61059	CCCTATGCTATCACATCCCGTGGTTTTCCCTTAGCTACTCAGTTGAATGTTCTCCTCCTCT	61118
Qy	439	CCCTCCTAGGGGTTTCCGGTTTGTCTCCTCTCTCAAGGTTTTTTTTCACAGCTGCAGGACC	498
Db	61119	CCCTCCTAGGGGTTTCCGGTTTGTCTCCTCTCTCAAGGTTTTTTTTCACAGCTGCAGGACC	61178
Qy	499	CGCTGCCCCACCTATTGCGAGCTGAGCCTGCTGCGAGCTGCACCTCTTACAGCCACACCTGT	558
Db	61179	CGCTGCCCCACCTATTGCGAGCTGAGCCTGCTGCGAGCTGCACCTCTTACAGCCACACCTGT	61238
Qy	559	AGCAGCTGAGCCTGCTGCAGGGGCCCTGTGTGACGCTGAGCCTGCTGCAGAGGACACCTGT	618
Db	61239	AGCAGCTGAGCCTGCTGCAGGGGCCCTGTGTGACGCTGAGCCTGCTGCAGAGGACACCTGT	61298
Qy	619	TGGAGCTGAGCCTGCTGCAGAGGCACCTGTGTGACGCTGAGCCTGCTGCAGAGGCACCTGT	678
Db	61299	TGGAGCTGAGCCTGCTGCAGAGGCACCTGTGTGACGCTGAGCCTGCTGCAGAGGCACCTGT	61358
Qy	679	TGGAGCTGGAGCCAGCTGCAGAGGAACCTTCCACCAGCTGAGCCTGCTACAGCCAAAGCCTGC	738
Db	61359	TGGAGCTGGAGCCAGCTGCAGAGGAACCTTCCACCAGCTGAGCCTGCTACAGCCAAAGCCTGC	61418
Qy	739	TGCCCCAGAACCTCACCTTCTCCCTCTCTTTGAACAGGCAAAATCAGTGAATTTCTTAGA	798
Db	61419	TGCCCCAGAACCTCACCTTCTCCCTCTCTTTGAACAGGTTAGTTTATATCTTAGCAC	61478
Qy	799	AGAGTACCATGGTTCAATTTCTATACTGATGCAGAAATAAAGTGAATCTACAAAAGTTTT	858
Db	61479	TATATGTATGAGAANTGAGATTGTTAGAGGGGGAAGAAAAAACCACTCCCCAAGCAA	61538
Qy	859	CTTCTTTTCCAAAGACTATTTCATTCCTGTTGTTATTTCAGA	898
Db	61539	ATCTATATCAACTTAATATTTTGAACAGCTTTATTTCAGA	61578
RESULT 3			
AC060228			
LOCUS	AC060228	157836 bp	DNA linear
DEFINITION	Homo sapiens chromosome 4 clone RP11-922D2, WORKING DRAFT SEQUENCE, 29 unordered pieces.		
ACCESSION	AC060228		
VERSION	AC060228.10	GI:20335630	
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT.		
SOURCE	Homo sapiens		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T., Barbarella,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Burch,C., Burch,P., Burket,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,		

Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Harris,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harrell,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogue,M., Holloway,C., Hollins,B., Homsli,F., Howard,S., Huber,J., Huiyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H., Lozados,R., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mahoney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M., Miner,G., Miner,T., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rivers,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoohtari,N., Sisson,I., Sodergren,E., Sonaik,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Wang,Q., Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Weinstock,G. and Gibbs,R.			
Direct Submission			
Unpublished			
2 (bases 1 to 157836)			
Worley,K.C.			
Direct Submission			
Submitted (20-APR-2000)	Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA		
3 (bases 1 to 157836)			
Worley,K.C.			
Direct Submission			
Submitted (09-MAY-2002)	Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA		
On Apr 28, 2002	this sequence version replaced gi:16117945.		
-----	Genome Center		
Center:	Baylor College of Medicine		
Center code:	BCM		
Web site:	http://www.hgsc.bcm.tmc.edu/		
Contact:	hgsc-help@bcm.tmc.edu		
-----	Project Information		
Center project name:	HBEB		
Center clone name:	RP11-922D2		
-----	Summary Statistics		
Sequencing vector:	M13;		
Chemistry:	Dye-primer	Bodypy: 93% of reads	
Chemistry:	Dye-terminator	Big Dye: 7% of reads	
Assembly program:	Phrap;	version 0.990329	
Consensus quality:	155736	bases at least Q40	
Consensus quality:	161898	bases at least Q30	
Consensus quality:	165232	bases at least Q20	
Estimated insert size:	166756;	sum-of-contigs estimation	
Quality coverage:	3x in Q20	bases; sum-of-contigs estimation	

* NOTE:	Estimated insert size may differ from sequence length (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).		
* NOTE:	This is a 'working draft' sequence. It currently consists of 29 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.		
* This record	will be updated with the finished sequence as soon as it is available and the accession number will be preserved.		
		2348: contig of 2348 bp in length	
		2448: gap of unknown length	
		1	
		2349	
		*	

Young,G., Zainoun,J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (17-MAY-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L7416
Center clone name: 653_L5

----- Summary Statistics
Sequencing vector: M13; M7815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 167163 bases at least Q40
Consensus quality: 175669 bases at least Q30
Consensus quality: 179699 bases at least Q20
Insert size: 188000; agarose-fp
Quality coverage: 3.6 in Q20 bases; agarose-fp
Quality coverage: 3.7 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 34 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1727: contig of 1727 bp in length
1728 1827: gap of 100 bp
1828 3423: contig of 1596 bp in length
3424 3523: gap of 100 bp
3524 6001: contig of 2478 bp in length
6002 6101: gap of 100 bp
6102 7881: contig of 1780 bp in length
7882 7981: gap of 100 bp
7982 10379: contig of 2398 bp in length
10380 10479: gap of 100 bp
10480 13080: contig of 2601 bp in length
13081 13180: gap of 100 bp
13181 13812: contig of 2632 bp in length
13813 15912: gap of 100 bp
15913 18292: contig of 2380 bp in length
18293 18392: gap of 100 bp
18393 20550: contig of 2158 bp in length
20551 20650: gap of 100 bp
20651 24116: contig of 3466 bp in length
24117 24216: gap of 100 bp
24217 27083: contig of 2867 bp in length
27084 27183: gap of 100 bp
27184 31082: contig of 3899 bp in length
31083 31182: gap of 100 bp
31183 36255: contig of 5073 bp in length
36256 36355: gap of 100 bp
36356 40430: contig of 4075 bp in length
40431 40530: gap of 100 bp
40531 44416: contig of 3886 bp in length
44417 44516: gap of 100 bp
44517 48353: contig of 3837 bp in length
48354 48453: gap of 100 bp
48454 52802: contig of 4349 bp in length
52803 52902: gap of 100 bp
52903 58991: contig of 6089 bp in length
58992 59091: gap of 100 bp
59092 63363: contig of 4272 bp in length
63364 63463: gap of 100 bp

* 63464 68027: contig of 4564 bp in length
* 68028 68127: gap of 100 bp
* 68128 72911: contig of 4784 bp in length
* 72912 73011: gap of 100 bp
* 73012 79146: contig of 6135 bp in length
* 79147 79246: gap of 100 bp
* 79247 86421: contig of 7175 bp in length
* 86422 86521: gap of 100 bp
* 86522 92006: contig of 5485 bp in length
* 92007 92106: gap of 100 bp
* 92107 97911: contig of 5805 bp in length
* 97912 98011: gap of 100 bp
* 98012 103109: contig of 5098 bp in length
* 103110 103209: gap of 100 bp
* 103210 109911: contig of 6702 bp in length
* 109912 110011: gap of 100 bp
* 110012 119453: contig of 9442 bp in length
* 119454 119553: gap of 100 bp
* 119554 126958: contig of 7405 bp in length
* 126959 127058: gap of 100 bp
* 127059 134408: contig of 7350 bp in length
* 134409 134508: gap of 100 bp
* 134509 143839: contig of 9331 bp in length
* 143840 143939: gap of 100 bp
* 143940 155329: contig of 11390 bp in length
* 155330 155429: gap of 100 bp
* 155430 168024: contig of 12595 bp in length
* 168025 168124: gap of 100 bp
* 168125 185969: contig of 17845 bp in length.

FEATURES

source
1..185969
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="4"
/map="4"
/clone="RP11-653L5"
/clone_lib="RPC1-11 Human Male BAC"
1..1727
/note="assembly_fragment"
1828..3423
/note="assembly_fragment"
3524..6001
/note="assembly_fragment"
6102..7881
/note="assembly_fragment"
7982..10379
/note="assembly_fragment"
10480..13080
/note="assembly_fragment"
13181..13812
/note="assembly_fragment"
15913..18292
/note="assembly_fragment"
18393..20550
/note="assembly_fragment"
20651..24116
/note="assembly_fragment"
24217..27083
/note="assembly_fragment"
clone_end:T7
vector_side:left
27184..31082
/note="assembly_fragment"
31183..36255
/note="assembly_fragment"
36356..40430
/note="assembly_fragment"
40531..44416
/note="assembly_fragment"
44517..48353
/note="assembly_fragment"
48454..52802
/note="assembly_fragment"

misc_feature		52903.. .58991	/note="assembly_fragment"	
misc_feature		59092.. .63363	/note="assembly_fragment"	
misc_feature		63464.. .68027	/note="assembly_fragment"	
misc_feature		68128.. .72911	/note="assembly_fragment"	
misc_feature		73012.. .79146	/note="assembly_fragment"	
misc_feature		79247.. .86421	/note="assembly_fragment"	
misc_feature		86522.. .92006	/note="assembly_fragment"	
misc_feature		92107.. .97911	/note="assembly_fragment"	
misc_feature		98012.. .103109	/note="assembly_fragment"	
misc_feature		103210.. .109911	/note="assembly_fragment"	
Query Match				
Best Local Similarity		53.3%;	Score 537.6;	DB 2; Length 185969;
Matches 610;		Conservative	0;	Mismatches 84; Indels 2; Gaps 2;
QY	203	GATGACAATGACGATGGTGACCCACTCATCCATCTCTGCAATCTCTGCAATTCCTTATGGCATACGG	262	
Db	167946	GATGACAATGACGATGGTGACCCACTCATCCATCTCTGCAATTCCTTATGGCATACGG	167887	
QY	263	AATTACACCTCTCTTTATATATCGCCAGTGAATACAGTCCCACTTACCTTGGGAAT	322	
Db	167886	AATTACACCTCTCTTTATATCGCCAGTGAATACAGTCCCACTTACCTTGGGAAT	167827	
QY	323	ACTTACACTGACAGGGTTACCTTCGTATCCCTGGATTCTTAATCTCTCGGATTCCTCC	382	
Db	167826	ACTTACACTGACAGGGTACCTTCGTATCCCTGGATTCTTAATCTCTCGGATTCCTCC	167767	
QY	383	TATGCTATACATCGTGGTTTCCCTTACTAGTCTACTCAGTTGAATGTCCTCTCCCT	442	
Db	167766	TATGCTATACATCGGAGGTTACCCCTTAATCTACTCAGTTGAATGTCCTCTCCCT	167707	
QY	443	CCTAGGGGTTTCCCGTTGTGTCCTTCCTCAAGGTTTTTTTTCAGCAGCTGCAGACCCGCT	502	
Db	167706	CCTA-GGGTTTCCGTTGTGTCCTTCCTTC-AGGTTTTTTTCAGCAGCTGCAGACCCGCT	167649	
QY	503	GCCCCACTATTGACGTGAGCCTGTCTGCAGCTGCACCTCTTACAGCCACACCTGTAGCA	562	
Db	167648	GCCCCACTATTGACGTGAGCCTGTCTGCAGCTGCACCTCTTACAGCCACACCTGTAGCA	167589	
QY	563	GCTGAGCCTGCTGAGGGGCCCTGTGTGACGTGAGCCTGCTGCAGAGGACCTGTTGGA	622	
Db	167588	GCTGAGCCTGCTGAGGGGCCCTGTGTGACGTGAGCCTGCTGCAGAGGACCTGTTGGA	167529	
QY	623	GCTGAGCCTGCTGAGAGGCACCTGTGTGACGTGAGCCTGCTGCAGAGGACCTGTTGGA	682	
Db	167528	GCTGAGCCTGCTGAGAGGCACCTGTGTGACGTGAGCCTGCTGCAGAGGACCTGTTGGA	167469	
QY	683	GTGAGGCCAGCTGCAGAGGAACCTTACCAGCTGAGCCTGCTACAGCCAAAGCCTGCTGCC	742	
Db	167468	GTGAGGCCAGCTGCAGAGGAACCTTACCAGCTGAGCCTGCTACAGCCAAAGCCTGCTGCC	167409	
QY	743	CCAGAACTCACCTCTCCCTCTCTTGACAGGCAATTCAGTGAATTCCTAGAAAGAG	802	
Db	167408	CCAGAACTCACCTCTCCCTCTCTTGACAGGTAAGTGGTTTATATCTTACCACATATA	167349	
QY	803	TACCATGGTTTCATTTCTATCTAGTCAGAAATAGTGAATCTACAAAAGTTTCTTT	862	
Db	167348	ATGTATGAGAATTCAGATTTGTAGAGGGGAAAAAACCACCTCCCAAGCAAAATCT	167289	
QY	863	CTTTTCCAAAGACTATTTCATTCCTTGTGTATTTCAGA	898	
Db	167288	ATAFCACTTAATATTTTGAACAGCTTTATTTCAGA	167253	
AC069037 Homo sapiens chromosome 4 clone RP11-653L5 map 4, WORKING DRAFT				
SEQUENCE, 34 unordered pieces.				
AC069037 HTG: HTGS_PHASE1; HTGS_DRAFT.				
Homo sapiens				
Homo sapiens				
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
1 (bases 1 to 185969)				
Birren,B., Linton,L., Nusbaum,C. and Lander,E.				
Unpublished				
2 (bases 1 to 185969)				
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N., Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F., Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N., Grant,G., Hacos,B., Heaford,A., Horton,L., Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J., LaRocque,K., Lamazares,K., Landers,T., Lehoczyk,J., Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrum,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Olivier,T.M., Oliver,J., Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Triglio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and Zody,M.				
Direct Submission				
Submitted (17-MAY-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA				
All repeats were identified using RepeatMasker:				
Smit, A.F.A. & Green, P. (1996-1997)				
http://ftp.genome.washington.edu/RM/RepeatMasker.html				
----- Genome Center				
Center: Whitehead Institute/ MIT Center for Genome Research				
Center code: WIBR				
Web site: http://www-seq.wi.mit.edu				
Contact: sequence_submissions@genome.wi.mit.edu				
----- Project Information				
Center project name: L74L6				
Center clone name: 653_L_5				
----- Summary Statistics				
Sequencing vector: M13; M7815; 100% of reads				
Chemistry: Dye-terminator Big Dye; 100% of reads				
Assembly program: Phrap; version 0.960731				
Consensus quality: 167163 bases at least Q40				
Consensus quality: 175669 bases at least Q30				
Consensus quality: 179699 bases at least Q20				
Insert size: 188000; agarose-fp				
Quality coverage: 3.6 in Q20 bases; agarose-fp				
Quality coverage: 3.7 in Q20 bases; sum-of-contigs				

* NOTE: This is a 'working draft' sequence. It currently				
* consists of 34 contigs. The true order of the pieces				
* is not known and their order in this sequence record is				
* arbitrary. Gaps between the contigs are represented as				
* runs of N, but the exact sizes of the gaps are unknown.				
* This record will be updated with the finished sequence				
* as soon as it is available and the accession number will				
* be preserved.				
* 1 1727: contig of 1727 bp in length				

*	1728	1827:	gap of	100 bp
*	1828	3423:	contig of	1596 bp in length
*	3424	3523:	gap of	100 bp
*	3524	6001:	contig of	2478 bp in length
*	6002	6101:	gap of	100 bp
*	6102	7881:	contig of	1780 bp in length
*	7882	7981:	gap of	100 bp
*	7982	10379:	contig of	2398 bp in length
*	10380	10479:	gap of	100 bp
*	10480	13080:	contig of	2601 bp in length
*	13081	13180:	gap of	100 bp
*	13181	15812:	contig of	2632 bp in length
*	15813	15912:	gap of	100 bp
*	15913	18292:	contig of	2380 bp in length
*	18293	18392:	gap of	100 bp
*	18393	20550:	contig of	2158 bp in length
*	20551	20650:	gap of	100 bp
*	20651	24116:	contig of	3466 bp in length
*	24117	24216:	gap of	100 bp
*	24217	27083:	contig of	2867 bp in length
*	27084	27183:	gap of	100 bp
*	27184	31082:	contig of	3899 bp in length
*	31083	31182:	gap of	100 bp
*	31183	36255:	contig of	5073 bp in length
*	36256	36355:	gap of	100 bp
*	36356	40430:	contig of	4075 bp in length
*	40431	40530:	gap of	100 bp
*	40531	44416:	contig of	3886 bp in length
*	44417	44516:	gap of	100 bp
*	44517	48353:	contig of	3837 bp in length
*	48354	48453:	gap of	100 bp
*	48454	52802:	contig of	4349 bp in length
*	52803	52902:	gap of	100 bp
*	52903	58991:	contig of	6089 bp in length
*	58992	59091:	gap of	100 bp
*	59092	63363:	contig of	4272 bp in length
*	63364	63463:	gap of	100 bp
*	63464	68037:	contig of	4564 bp in length
*	68028	68127:	gap of	100 bp
*	68128	72911:	contig of	4784 bp in length
*	72912	73011:	gap of	100 bp
*	73012	79146:	contig of	6135 bp in length
*	79147	79246:	gap of	100 bp
*	79247	86421:	contig of	7175 bp in length
*	86422	86521:	gap of	100 bp
*	86522	92006:	contig of	5485 bp in length
*	92007	92106:	gap of	100 bp
*	92107	97911:	contig of	5805 bp in length
*	97912	98011:	gap of	100 bp
*	98012	103109:	contig of	5098 bp in length
*	103110	103209:	gap of	100 bp
*	103210	109911:	contig of	6702 bp in length
*	109912	110011:	gap of	100 bp
*	110012	119453:	contig of	9442 bp in length
*	119454	119553:	gap of	100 bp
*	119554	126958:	contig of	7405 bp in length
*	126959	127058:	gap of	100 bp
*	127059	134408:	contig of	7350 bp in length
*	134409	134508:	gap of	100 bp
*	134509	143839:	contig of	9331 bp in length
*	143840	143939:	gap of	100 bp
*	143940	155329:	contig of	11390 bp in length
*	155330	155429:	gap of	100 bp
*	155430	168024:	contig of	12595 bp in length
*	168025	168124:	gap of	100 bp
*	168125	185969:	contig of	17845 bp in length.
Location/Qualifiers				
1. .185969				
/organism="Homo sapiens"				
/db_xref="taxon:9606"				
/chromosome="4"				
/map="4"				
/clone="RP11-6531.5"				

QY	379	CCCTATGCTATACATCCCGTGGTTTCCCTTACGCTACTCAGTTGAATGTTCCCTCT	438
Db	67787	CCCTATGCTATACATCCCGTGGTTTCCCTTACGCTACTCAGTTGAATGTTCCCTCT	67846
QY	439	CCCTCCATAGGGTTTCCGTTTGTGTCCTTCAAGGTTTTCACAGCTGCAGCACC	498
Db	67847	CCCTCCATAGGGTTTCCGTTTGTGTCCTTCAAGGTTTTCACAGCTGCAGCACC	67906
QY	499	CGTGCCCCACCTATTGCAGCTGCAGCTGCTGCAGCTGCACCTCTTACAGCCACACCTGT	558
Db	67907	CGTGCCCCACCTATTGCAGCTGCAGCTGCTGCAGCTGCACCTCTTACAGCCACACCTGT	67966
QY	559	AGCAGCTGAGCCTGCTGCAGGGGGCCCTGTGTGCAGCTGAGCCTGCTGCAGGACCTGT	618
Db	67967	AGCAGCTGAGCCTGCTGCAGGGGGCCCTGTGTGCAGCTGAGCCTGCTGCAGGACCTGT	68026
QY	619	T 619	
Db	68027	T 68027	
RESULT 6			
AC099443/c			
LOCUS			
DEFINITION	AC099443	177035 bp DNA linear	HTG 12-JUL-2002
ACCESSION	AC099443		
VERSION	AC099443.3	GI:21729719	
KEYWORDS	HTG; HTGS_PHASE1.		
SOURCE	Norway rat.		
ORGANISM	Rattus norvegicus		
REFERENCE			
AUTHORS	Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Alsbrooks,S.L., Anaratunge,H.C., Are,J.R., Ayele,M., Banks,T., Barbaria,J., Benton,J., Bimaye,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carton,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gortell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Homsí,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louissege,H., Lozard,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newton,S., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoohtari,N., Sisson,I., Sodergren,E., Sonaite,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczkyk,R., Wooden,S., Woreley,K.,		

QY	533	GCTGCACCTCTTACAGCCACACCTGTAGCAGCTGAGCCTGCTGTCAGGGGCCCTGTTGCCA	592
Db	19929	GCTCATCCTGCTGCTCATCTGCTGCTCATCTGCTGCTGCTCATCTGCTGCTCATCTGCTGCT	19988
QY	593	GCTGAGCTGCTGCAGAGGACCTGTTGGAGCTGAGCCTGCTGCAGAGGACCTGTTGCCA	652
Db	19989	GCTCATCCTGCTGCTGCTCATCTGCTGCTGCTCATCTGCTGCTGCTGCTCATCTGCTGCT	20048
QY	653	GCTGAGCCTGCTGCAGAGGACCTGTTGGAGCTGAGCCAGCTGCAGAGGACCTTCACCA	712
Db	20049	GCTCATCCTGCTGCTGCTCATCTGCTGCTGCTCATCTGCTGCTGCTCATCTGCTGCTGCT	20108
QY	713	GCTGAGCCTGCTACAGCAAGCCTGCTGCCCCAGAACCTCACCCCTTCCCTCTCTTG	770
Db	20109	GCTCATCCTGCTGCTGCTCATCTGCTGCTGCTCATCTGCTGCTGCTGCTCATCTGCTGCTG	20166
RESULT	13		
KSU75698			
LOCUS		137508 bp	DNA linear VRL 03-MAY-1997
DEFINITION		Kaposi's sarcoma-associated herpesvirus long unique region, 80 putative ORF's and kaposin gene, complete cds.	
ACCESSION		U75698	
VERSION			
KEYWORDS		U75698.1	GI:2065526
SOURCE			
ORGANISM		Human herpesvirus 8.	
		Human herpesvirus 8	
		Viruses; dsDNA viruses, no RNA stage; Herpesviridae; Gammaherpesvirinae; Rhadinovirus.	
REFERENCE		1 (bases 47193 to 47522; 133099 to 133729)	
AUTHORS		Chang,Y., Cesarman,E., Pessin,M.S., Lee,F., Culpepper,J., Knowles,D.M. and Moore,P.S.	
TITLE		Identification of herpesvirus-like DNA sequences in AIDS-associated Kaposi's sarcoma	
JOURNAL		Science 266 (5192), 1865-1869 (1994)	
MEDLINE		95090463	
PUBMED		7997879	
REFERENCE		2 (bases 35021 to 55726)	
AUTHORS		Moore,P.S., Gao,S.J., DomInquez,G., Cesarman,E., Lungu,O., Knowles,D.M., Garber,R., Pellett,P.E., McGeoch,D.J. and Chang,Y.	
TITLE		Primary characterization of a herpesvirus agent associated with Kaposi's sarcoma	
JOURNAL		J. Virol. 70 (1), 549-558 (1996)	
MEDLINE		96099469	
PUBMED		8523568	
REFERENCE		3 (bases 28661 to 29741; 117919 to 118101)	
AUTHORS		Zhong,W., Wang,H., Herndier,B. and Ganem,D.	
TITLE		Restricted expression of Kaposi sarcoma-associated herpesvirus (human herpesvirus 8) genes in Kaposi sarcoma	
JOURNAL		Proc. Natl. Acad. Sci. U.S.A. 93 (13), 6641-6646 (1996)	
MEDLINE		96270595	
PUBMED		8692871	
REFERENCE		4 (bases 122794 to 123567; 123809 to 127297; 129372 to 130400; 130551 to 134441)	
AUTHORS		Cesarman,E., Nador,R.G., Bai,F., Bohenzky,R.A., Russo,J.J., Moore,P.S., Chang,Y. and Knowles,D.M.	
TITLE		Kaposi's sarcoma-associated herpesvirus contains G protein-coupled receptor and cyclin D homologs which are expressed in Kaposi's sarcoma and malignant lymphoma	
JOURNAL		J. Virol. 70 (11), 8218-8223 (1996)	
MEDLINE		97048116	
PUBMED		8892957	
REFERENCE		5 (bases 1 to 137508)	
AUTHORS		Moore,P.S., Boshoff,C., Weiss,R.A. and Chang,Y.	
TITLE		Molecular mimicry of human cytokine and cytokine response pathway genes by KSHV	
JOURNAL		Science 274 (5293), 1739-1744 (1996)	
MEDLINE		97094384	
PUBMED		8939871	
REFERENCE		6 (bases 1 to 137508)	
AUTHORS		Russo,J.J., Bohenzky,R.A., Chien,M.-C., Chen,J., Yan,M., Maddalena,D., Parry,J.P., Peruzzi,D., Edelman,I.S., Chang,Y. and Moore,P.S.	

TITLE	Nucleotide sequence of the Kaposi sarcoma-associated herpesvirus (HHV8)		
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 93 (25), 14862-14867 (1996)		
MEDLINE	97121480		
PUBMED	8962146		
REFERENCE	7 (bases 1 to 137508)		
AUTHORS	Russo,J.J., Bohenzky,R.A., Chien,M.-C., Chen,J., Yan,M., Maddalena,D., Parry,J.P., Peruzzi,D., Edelman,I.S., Chang,Y. and Moore,P.S.		
TITLE	Direct Submission		
JOURNAL	Submitted (17-OCT-1996) Dept of Pathology/Genome Center/Division of Epidemiology, Columbia University, 650 West 168th Street, New York, NY 10032, USA		
REFERENCE	8 (bases 1 to 137508)		
AUTHORS	Russo,J.J., Bohenzky,R.A., Chien,M.-C., Chen,J., Yan,M., Maddalena,D., Parry,J.P., Peruzzi,D., Edelman,I.S., Chang,Y. and Moore,P.S.		
TITLE	Direct Submission		
JOURNAL	Submitted (02-MAY-1997) Dept of Pathology/Genome Center/Division of Epidemiology, Columbia University, 650 West 168th Street, New York, NY 10032, USA		
REMARK	Sequence update by submitter		
COMMENT	On May 3, 1997 this sequence version replaced gi:1718251.		
FEATURES	Location/Qualifiers		
source	1..137508		
	/organism="Human herpesvirus 8"		
	/db_xref="taxon:37296"		
	/cell_line="BC-1; virally infected primary effusion lymphoma derived cell line"		
	/note="sequence derived from overlapping lambda phage and cosmid clones isolated from libraries; A prototype sequence for the terminal repeat (TR) unit is deposited under GenBank Accession Number U75699. The left junction variant TR sequence unit immediately preceding base pair 1 of the LUR is deposited under GenBank Accession Number U75700. The remaining LUR sequence extending from this sequence to the right TR unit is not available due to difficulties in cloning and sequencing of this region"		
misc_feature	1..137508		
	/note="long unique region LUR"		
CDS	105..974		
	/note="ORF K1"		
	/codon_start=1		
	/protein_id="AAC57081.1"		
	/db_xref="GI:1718252"		
	/translation="MFLYVCSLAVCFRGLLSLLSPNLCPGVISTPYTLCLSNASLPTISWCNTRLLRLTERRVILDTIACNFTCEQSHRQSIWITWRAQPVLIQLCAQPSNITVCQHVTLYCSTSGNNVTVMHLPNGRNETVSOIKYNTLMSTGECYCSNGLSSRSLNRICFWARCANITPETHTVSVSSTTGFTLSLTSNLVKIITHATRDVVVKEAKCTHPHIEVHFLVMTLVALLICTMCGILGTIIFAHCQQRDSNTVPQQLQDYISLHDKTDEYTOPDWY"		
CDS	1142..2794		
	/note="ORF 4; The HVS ORF 4 homolog has alternatively spliced messages encoding membrane bound and secreted proteins; complement binding protein vCBP homolog"		
	/codon_start=1		
	/protein_id="AAC57082.1"		
	/db_xref="GI:1718253"		
	/translation="MAFLRQLTWLWTFMTWIGQDNEKCSOKTLIGYRLKMSRDGDIAGVETVELRCRGYTTIARNITATCIQGTWSEPTATCNKSCPNPGQTQNGKVFHFGQDALKYGANISYVCGNEGYFLVGRVYRYCMASGQMAWSSSPFCEKEKVRPKIKNGDFKPKDYVEYNDVAHPECNEGTVLVGPHSIACAVANNWTSMPTCELACGKFPSTVHGYPLOGESLTYKHQSQVTFACNDGFVLRGSPFTTCNTVEMDPLPKCVLEDIDDPNN SNPGLHTPNEKPNGVNFQSNTEPPTPKPEDTHATATCNCQPKVLPTSEGN DATTSNITIKOLEDEKTLISQNTIHITSALTSMKAGNFTNKNSTLDLHASTPTSD ETTIPSVQTPNTNAPTRTLASLHIEGSPNSTSEKATSTLSLHNSKNDTGGI YTTLNKTLQLPSTNKPNSQAKSTKPRVETHNKTISNPALSLTSDSADVPQRPREPTLPPIERPAPSKNRYLEKQLVIGLLTAVALTCGLITFLHYLFFR"		
CDS	3210..6611		
	/note="ORF 6; ss DNA binding protein ssDBP homolog; EBV BALF2 homolog"		
	/codon_start=1		
	/protein_id="AAC57083.1"		


```
gene          14111..15262
CDS           /gene="ORF 74"
              14111..15139
              /gene="ORF 74"
              /note="G protein coupled receptor; similar to IL-8
              receptor; transforming protein"
              /codon_start=1
              /evidence=experimental
              /product="v-GPCR"
              /protein_id="A4D4503.1"
              /db_xref="GI:5669896"
              /translation="MAAEPLFIIFLDDSNWNETLMSGYDSGNFSLVSCMTTV
              VPTWNVGLSLIFLVLNGLVYIFCKHRSRAGADILLGLGICLSLSTSLA

Query Match      11.7%; Score 117.8; DB 14; Length 28559;
Best Local Similarity 50.6%; Pred. No. 1.2e-13;
Matches 284; Conservative 0; Mismatches 277; Indels 0; Gaps 0;

QY 217 TGGTCACCCACTTCATCATCTCGAATATTCCTTATGGCATACGGAATTTACCACTCC 276
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 9873 TGTCTCTCTGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 9932

QY 277 TCTTTATTATGCCAGTGAATACAGTCCCAAGTTACCTCGGGAATACTTACACTGACAC 336
      || || || || || || || || || || || || || || || || || || || ||
Db 9933 TCTTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 9992

QY 337 AGGTTTACCTTCGTATCCCTGGATTCTAACTCTCTCGGATTCCTCTATGCTATCATCAT 396
      || || || || || || || || || || || || || || || || || || || ||
Db 9993 TCTTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 10052

QY 397 CCGTGGTTTTCCCTTAGCTACTAGTTGAATGTTCCTCTCTCTCTCTCTCTCTCTCTCTCT 456
      || || || || || || || || || || || || || || || || || || || ||
Db 10053 TGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 10112

QY 457 GTTTGTCCCTCCCTTCAAGTTTTTTTTCAGCAGCTGCAGCAGCCGCTGCCCCACCTATTGC 516
      || || || || || || || || || || || || || || || || || || || ||
Db 10113 TCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 10172

QY 517 AGCTGAGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 576
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 10173 TGTCTATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 10232

QY 577 AGGGCCCTCTGTTGAGCTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 636
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 10233 TGTCTATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 10292

QY 637 AGAGCCACTGTTGAGCTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 696
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 10293 TGTGGCTCTCTGCTGTTGTGGCTCCTGCTGTTGTGGCTCCTGCTGTTGTGGCTCCTGCTG 10352

QY 697 AGAGGAACCTTCACAGCTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 756
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 10353 TGTGGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 10412

QY 757 TTCTCCCTCTCTTGAACAGGC 777
      || || || || || || || || || || || || || || || || || || || ||
Db 10413 TGTGGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 10433
```

RESULT 15

```
LOCUS      U93872      133661 bp      DNA      linear      VRL 09-JUL-2001
DEFINITION Kaposi's sarcoma-associated herpesvirus glycoprotein M, DNA
            replication protein, glycoprotein, DNA replication protein, FLICE
            inhibitory protein and v-cyclin genes, complete cds, and tegument
            protein gene, partial cds.
ACCESSION  U93872
VERSION    U93872.2  GI:14627174
KEYWORDS   Human herpesvirus 8.
SOURCE     Human herpesvirus 8.
            Viruses; dsDNA viruses, no RNA stage: Herpesviridae;
            Gammaherpesvirinae; Rhadinovirus.

REFERENCE  1 (bases 29032 to 30108; 117733 to 118431)
            Zhong,W., Wang,H., Herndier,B. and Ganem,D.
            Restricted expression of Kaposi sarcoma-associated herpesvirus
            (human herpesvirus 8) genes in Kaposi sarcoma
            Proc. Natl. Acad. Sci. U.S.A. 93 (13), 6641-6646 (1996)
            96270595
            2 (bases 17242 to 17856)
            Neipel,F., Albrecht,J.-C., Ensser,A., Huang,Y.-Q., Li,J.J.,
            Friedmann-Kien,A.E. and Fleckenstein,B.
            Human herpesvirus 8 encodes a homolog of interleukin-6
            J Virol. 71 (1), 839-842 (1997)
            97138401
            3 (bases 123309 to 124082)
            Li,M., Lee,H., Yoon,D.W., Albrecht,J.C., Fleckenstein,B., Neipel,F.
            and Jung,J.U.
            Kaposi's sarcoma-associated herpesvirus encodes a functional cyclin
            J. Virol. 71 (3), 1984-1991 (1997)
            97184528
            4 (bases 122660 to 123226)
            Thome,M., Schneider,P., Hofmann,K., Fickenscher,H., Meinel,E.,
            Neipel,F., Mettmann,C., Burns,K., Bodmer,J.L., Schroter,M.,
            Scarfidi,C., Krammer,P.H., Peter,M.E. and Tschopp,J.
            Viral FLICE-inhibitory proteins (FLIPs) prevent apoptosis induced
            by death receptors
            Nature 386 (6624), 517-521 (1997)
            97242415
            5 (bases 1 to 133661)
            Neipel,F., Albrecht,J.C. and Fleckenstein,B.
            Cell-homologous genes in the Kaposi's sarcoma-associated
            rhadinovirus human herpesvirus 8: determinants of its
            pathogenicity?
            J. Virol. 71 (6), 4187-4192 (1997)
            97296220
            6 (bases 1 to 133661)
            Neipel,F., Albrecht,J.-C., Ensser,A., Huang,Y.-Q., Li,J.J.,
            Friedmann-Kien,A.E. and Fleckenstein,B.
            The genome of human herpesvirus 8 cloned from Kaposi's sarcoma
            Unpublished
            7 (bases 1 to 133661)
            Neipel,F., Albrecht,J.-C., Ensser,A., Huang,Y.-Q., Li,J.J.,
            Friedmann-Kien,A.E. and Fleckenstein,B.
            Direct Submission
            Submitted (17-MAR-1997) Virology, University of Erlangen,
            Schlossgarten 4, Institut für Klinische und Molekulare, Erlangen
            91054, Germany
            8 (bases 1 to 133661)
            Neipel,F., Albrecht,J.-C., Ensser,A., Huang,Y.-Q., Li,J.J.,
            Friedmann-Kien,A.E. and Fleckenstein,B.
            Direct Submission
            Submitted (09-JUL-2001) Virology, University of Erlangen,
            Schlossgarten 4, Institut für Klinische und Molekulare, Erlangen
            91054, Germany
            REMARK      Sequence update by submitter
            COMMENT      On Jul 9, 2001 this sequence version replaced gi:2246466.
            FEATURES
            source
            1..133661
            /organism="Human herpesvirus 8"
            /db_xref="taxon:37296"
            /note="sequence derived from overlapping Lambda phage
            clones isolated from two libraries of Sau3A partially
            digested DNA from Kaposi's sarcoma biopsies"
            105..959
            /note="ORF K1"
            /codon_start=1
            /protein_id="AAB62619.1"
            /db_xref="GI:2246494"
            /translation="MFLYVVCSLAVCFRGLLSLSQSSPNLCPCGVISTPYTTLTQPSNT
            SLPTSWYCNDRLLRLTQQTFTVDTLICNFCVCGSGHRSLSLWYWPQVLTLCGQ
```


Job time : 3379.23 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model
Run on: February 10, 2003, 11:43:03 ; Search time 256.086 Seconds
(without alignments)
8864.249 Million cell updates/sec

Title: US-09-923-236-1
Perfect score: 1008
Sequence: 1 agacagactaaaaagccat.....tagataatttagaccaatgg 1008

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DE seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_101002.*
1: /SID22/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SID22/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SID22/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SID22/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SID22/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SID22/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SID22/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SID22/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SID22/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SID22/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SID22/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SID22/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SID22/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SID22/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SID22/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SID22/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SID22/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SID22/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SID22/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SID22/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SID22/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SID22/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SID22/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SID22/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1008	100.0	1008	24	Human zsig63 cDNA.
2	997	98.9	998	22	Human transport pr
3	981.6	97.4	1325	20	Human secreted pro
4	896.4	88.9	959	20	DNA encoding a hum
5	460	45.6	657	24	Human zsig63 degen
6	449	44.5	470	20	EST clone CP251.
7	421	41.8	450	20	EST clone CP116.
8	286.4	28.4	296	20	EST clone CP328.
C	121.2	12.0	3489	21	Kaposi's sarcoma-a

C	10	121.2	12.0	3489	22	AAF82901	Nucleotide sequenc
C	11	121.2	12.0	3489	24	ABA93487	Kaposi's sarcoma-a
C	12	121.2	12.0	32207	20	AAV73805	KSHV LUR DNA (nucl
C	13	121.2	12.0	137507	19	AAV19941	KSHV long unique c
C	14	106	10.5	5120	22	AAAC84677	DNA sequence of hu
C	15	100	9.9	1037	21	AAAS9242	Exons E, C and A o
C	16	100	9.9	1472	21	AAAS9241	Exons D, C, B and
C	17	99.2	9.8	1159	21	AAAS9240	An EcoRI fragment
C	18	98.2	9.7	2108	24	ABL67774	Oesophagus cancer
C	19	93.6	9.3	397	20	AAK89891	Spinocerebellar at
C	20	92.8	9.2	5120	22	AAAC84677	DNA sequence of hu
C	21	92.4	9.2	1101	22	AAAD06303	Human cDNA clone H
C	22	92	9.1	543	13	AAQ23092	Antigen tc-7a gene
C	23	89.8	8.9	3042	23	ABL28445	Drosophila melanog
C	24	89.8	8.9	5215	23	ABL28444	Drosophila melanog
C	25	88.4	8.8	403	21	ABK10240	Trinucleotide repe
C	26	88.2	8.8	627	23	AAAS93555	DNA encoding novel
C	27	87.6	8.7	234	16	AAQ84832	Spinocerebellar at
C	28	87.6	8.7	1101	22	AAAD06303	Human cDNA clone H
C	29	86.6	8.6	786	17	AAAT46148	T. cruzi L19E homo
C	30	86.6	8.6	786	20	AAZ10978	TcE coding sequenc
C	31	86.4	8.6	486	22	AAF75507	Polyglutamine trac
C	32	85.2	8.5	799	19	AAV55831	Nucleotide sequenc
C	33	85.2	8.5	1926	21	AAAS0254	Epstein Barr virus
C	34	85.2	8.5	1926	22	AAAF82902	EBV tethering prot
C	35	85.2	8.5	2580	21	AAAF75454	Nucleotide sequenc
C	36	85.2	8.5	2580	24	AAAF64275	Epstein-Barr virus
C	37	85.2	8.5	5452	20	AAAG09923	Anti-sense strand
C	38	85.2	8.5	8705	20	AAZ23778	Vector pshuttle DN
C	39	85.2	8.5	9600	19	AAV21683	Vector plasmid pCM
C	40	85.2	8.5	10380	20	AAZ22248	Nucleotide sequenc
C	41	85.2	8.5	10596	14	AAQ51731	Plasmid pCisEBON f
C	42	85.2	8.5	10596	17	AAAT40348	Plasmid pCisEBON f
C	43	85.2	8.5	10596	20	AAAX15650	Nucleotide sequenc
C	44	85.2	8.5	16080	21	AAAS9553	DNA clone pCEK Cl.
C	45	84	8.3	795	19	AAV55830	FLGA insert stabil

ALIGNMENTS

RESULT 1	AAAS20591	ID	AAAS20591 standard; cDNA; 1008 Bp.
XX	AAAS20591;	AC	AAAS20591;
XX	23-APR-2002	DT	(first entry)
XX	Human zsig63 cDNA.	DE	Human zsig63 cDNA.
XX	Human; zsig63; chromosome 4q12-4q13; salivary protein; antimicrobial; ss;	KW	Human; zsig63; chromosome 4q12-4q13; salivary protein; antimicrobial; ss;
KW	microbial infection; tooth decay; periodontal disease; thrush; emphysema;	KW	microbial infection; tooth decay; periodontal disease; thrush; emphysema;
KW	gastrointestinal disease; urinary tract infection; vaginal infection;	KW	gastrointestinal disease; urinary tract infection; vaginal infection;
KW	skin infection; epithelial wound; chronic tissue damage; cystic fibrosis;	KW	skin infection; epithelial wound; chronic tissue damage; cystic fibrosis;
KW	acquired immunodeficiency syndrome; AIDS; lung infection; sarcoidosis;	KW	acquired immunodeficiency syndrome; AIDS; lung infection; sarcoidosis;
XX	chronic bronchitis; gene therapy; protein therapy; gene.	XX	chronic bronchitis; gene therapy; protein therapy; gene.
OS	Homo sapiens.	OS	Homo sapiens.
XX	Key	FH	Location/Qualifiers
FT	128..787	FT	128..787
FT	/*tag= a	FT	/*tag= a
FT	/product= "Human zsig63"	FT	/product= "Human zsig63"
XX	US6331413-B1.	PN	US6331413-B1.
XX	18-DEC-2001.	PD	18-DEC-2001.
XX	17-MAR-2000; 2000US-0527345.	PF	17-MAR-2000; 2000US-0527345.
XX	17-MAR-1999; 99US-124820P.	XX	17-MAR-1999; 99US-124820P.
XX		XX	

SQ Sequence 998 BP; 253 A; 249 C; 190 G; 306 T; 0 other;

Query Match 98.9%; Score 997; DB 22; Length 998;
Best Local Similarity 100.0%; Pred. No. 3.3e-203;
Matches 997; Conservative 0; Mismatches 0; Indels

0.

Qy	6	GACTAAAGAAGCCATGTATCTTTCGGTTCTCTCPTAAAAGAGAATAAATTAAAA	65
Db	1	GACTAAAAGGCCATGTATCTTTCGGTTCTCTCPTAAAAGAGAATAAATTAAAA	60
Qy	66	ATACATTGCGTATTTCTCTAAAACAATAAATTAATAGTGTAATATTCATAGGGTCAATCA	125
Db	61	ATACATTGCGTATTTCTCTAAAACAATAAATTAATAGTGTAATATTCATAGGGTCAATCA	120
Qy	126	AAATGAAGCTTCTCCTTTGGGCCTGCAATGTATGTGTCTGCTTTTCCAAGGAAGACAGCGT	185
Db	121	AAATGAAGCTTCTCCTTTGGGCCTGCAATGTATGTGTCTGCTTTTCCAAGGAAGACAGCGT	180
Qy	186	TCCCCCTTCAATTGGTCAGGATGACAATGACCATGGTCACCACCTTCATCCATCTCTGGAATA	245
Db	181	TCCCCCTTCAATTGGTCAGGATGACAATGACCATGGTCACCACCTTCATCCATCTCTGGAATA	240
Qy	246	TTCCTTATGGGATACGGAAATTTACACCTCTCTTTATTTATCGCCCAGTGAATACAGTCC	305
Db	241	TTCCTTATGGGATACGGAAATTTACACCTCTCTTTATTTATCGCCCAGTGAATACAGTCC	300
Qy	306	CCAGTTTACCCCTGGGAATACITTTACACTGACACAGGGTTTACCTTCGTATCCCTGGATTCCTAA	365
Db	301	CCAGTTTACCCCTGGGAATACITTTACACTGACACAGGGTTTACCTTCGTATCCCTGGATTCCTAA	360
Qy	366	CTTCTCCTCGATTCGCCCTATGTCTATCACATPCGTGGTTTTCCCTTAGCTACTCAGTTGA	425
Db	361	CTTCTCCTCGATTCGCCCTATGTCTATCACATPCGTGGTTTTCCCTTAGCTACTCAGTTGA	420
Qy	426	ATGTTCTCCTCCTCCTCTCTAGGGGTTTCCCGTTTGTCCCTCTCTCAAAGGTTTTTTTTTCAG	485
Db	421	ATGTTCTCCTCCTCCTCTCTAGGGGTTTCCCGTTTGTCCCTCTCTCAAAGGTTTTTTTTTCAG	480
Qy	486	CAGCTGCAGACCCCGCTGCCCACTATTTGCAGCTGAGCGCTGCTGCAGCTGCACCTCTTA	545
Db	481	CAGCTGCAGACCCCGCTGCCCACTATTTGCAGCTGAGCGCTGCTGCAGCTGCACCTCTTA	540
Qy	546	CAGGCACACCTGTAGCAGCTGAGCGCTGCTGCAGGGGCCCTGTTGCAGCTGAGCGCTGCTG	605
Db	541	CAGGCACACCTGTAGCAGCTGAGCGCTGCTGCAGGGGCCCTGTTGCAGCTGAGCGCTGCTG	600
Qy	606	CAGAGCGACCTGTTGGAGCTGAGCGCTGCTGCAGAGGCACTGTTGGAGCTGAGCGCTGCTG	665
Db	601	CAGAGCGACCTGTTGGAGCTGAGCGCTGCTGCAGAGGCACTGTTGGAGCTGAGCGCTGCTG	660
Qy	666	CAGAGCGACCTGTTGGAGTGAGCGCAGCTGCAGAGGAACCTTCACAGCTGAGCGCTGCTA	725
Db	661	CAGAGCGACCTGTTGGAGTGAGCGCAGCTGCAGAGGAACCTTCACAGCTGAGCGCTGCTA	720
Qy	726	CAGCAAGCGCTGCTGCCCGCAGAACCTCACCCCTCTCTCCCTCTTGTGAACAGGCAATCAGT	785
Db	721	CAGCAAGCGCTGCTGCCCGCAGAACCTCACCCCTCTCTCCCTCTTGTGAACAGGCAATCAGT	780
Qy	786	GAAATTTCTCTAGAAGTACCATGGGTTTCATTTCTATCTGATGCAGAAATAAGTGAAT	845
Db	781	GAAATTTCTCTAGAAGTACCATGGGTTTCATTTCTATCTGATGCAGAAATAAGTGAAT	840
Qy	846	CTACAAAAGTTTCTTTCTTTTCCAAGACTATTTTCATCTGTTGTATTCAGAGTATFCA	905
Db	841	CTACAAAAGTTTCTTTCTTTTCCAAGACTATTTTCATCTGTTGTATTCAGAGTATFCA	900
Qy	906	TCTCACTACATTGATTGTTTGTGTAGTTTTTCCCTTGACTTAATTTATATTCGAAAAA	965
Db	901	TCTCACTACATTGATTGTTTGTGTAGTTTTTCCCTTGACTTAATTTATATTCGAAAAA	960
Qy	966	CATTGATAAATTAATAAATAAATAGATAAATTTAGAC	1002
Db	961	CATTGATAAATTAATAAATAAATAGATAAATTTAGAC	997

RESULT 3

AAV80740
ID AAV80740 standard; cDNA; 1325 BP.

AAV80740;

DT 12-APR-1999 (first entry)

DE Human secreted protein clone cp116 1 encoding cDNA.

Human; secreted protein; nutritional activity; cytokine; vaccine; cell proliferation; differentiation; immune stimulation; suppression; haematopoiesis regulation; tissue growth; activin; inhibitor; chemotactic; chemokine; haemostatic; thrombolytic; anti-inflammatory; gene therapy; tumour invasion suppression; tumour inhibition; ds.

OS Homo sapiens.

XX	Key	Location/Qualifiers
FH	CDS	127..786
FT		

FT /product= "secreted protein clone cp116 1"

PN WO9901466-A1.

14-JAN-1999.

PF 01-JUL-1998; 98WO-US13813.

PR 27-OCT-1997; 97US-0958304.

XX
XX
CCT/800-8016, 1667 700-70 43

PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Racie LA;
PI Spaulding V, Treacy M;

DR WPI; 1999-105994/09.

XX New polynucleotides encoding secreted human proteins - are derived
PT from human foetal brain, adult testes, adult brain, foetal kidney,
PT adult salivary gland, or adult blood cDNA libraries, useful as, e.g.
PT potential vaccines

PS Claim 23; Page 70-71; 107pp; English.

The present sequence encodes a human secreted protein from clone csp16 1, deposited as ATCC 99482. Human secreted protein clone polynucleotides and proteins are predicted to have biological activities which would make them suitable for treating, preventing or ameliorating medical conditions in humans and animals. Suggested activities include nutritional activity, cytokine and cell proliferation/differentiation activity, immune stimulating (e.g. as vaccines) or suppressing activity, haematopoiesis regulating activity, tissue growth activity, activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic and thrombolytic activity, receptor/ligand activity, anti-inflammatory activity, cadherin/tumour invasion suppressor activity, and tumour inhibition activity. The polynucleotides are also stated to be useful for gene therapy.

Sequence 1325 BP; 407 A; 289 C; 241 G; 388 T; 0 other;

Query Match 97.4%; Score 981.6; DB 20; Length 1325;
Best Local Similarity 99.5%; Pred. No. 7e-200;

Matches 995; Conservative 0; Mismatches 4; Indels 1; Gaps 1;

QY 10 AAAAAGCCATGTATCTTTTCGTTCTCTAAAGAGAAAATATAATTAAAAATAC 69

Db 9 AAAAAGCCATGTATCTTTCGTTCTCTAAAAGAAGAAAAATATAATTAAAAATAC 68

Qy 70 ATTGCGTATTTTCTAAACAAATAAAATTTATAGTGTAAATATTCATAGGGTCAATCAAAAT 129
Dy 69 ATTGCGTATTTTCTAAACAAATAAAATTTATAGTGTAAATATTCATAGGGTCAATCAAAAT 128
Qy 130 GAAGCTTCTCTTGGGCGCTGCATTTATGTGTGCTTTTGAAGGAAGACGGTTCTCC 189
Dy 129 GAAGCTTCTCTTGGGCGCTGCATTTATGTGTGCTTTTGAAGGAAGACGGTTCTCC 188
Qy 190 CTTCAATGGTGAGGATGACAAATGACGATGGTGACCCACTTCATCCTGAATATTTCC 249
Dy 189 CTTCAATGGTGAGGATGACAAATGACGATGGTGACCCACTTCATCCTGAATATTTCC 248
Qy 250 TTATGGCATACGGAATTTACCACTCCCTCTTATTAATCGCCAGTGAATACAGTCCCCAG 309
Dy 249 TTATGGCATACGGAATTTACCACTCCCTCTTATTAATCGCCAGTGAATACAGTCCCCAG 308
Qy 310 TTACCCCTGGGAATCTTACACTGACACAGGGTTACCTTCGTATCCCTGGATTTCACTTC 369
Dy 309 TTACCCCTGGGAATCTTACACTGACACAGGGTTACCTTCGTATCCCTGGATTTCACTTC 368
Qy 370 TCCCTGGATTCCTCTATGCTATCACATCCGTGGTTTCCCTTAGCTACTCAGTTGAATGT 429
Dy 369 TCCCTGGATTCCTCTATGCTATCACATCCGTGGTTTCCCTTAGCTACTCAGTTGAATGT 428
Qy 430 TCCTCTCTCCCTCCTAGGGGTTTCCCGTTTGGTCCCTCTTCAAGGTTTTTTTTCAGCAGC 489
Dy 429 TCCTCTCTCCCTCCTAGGGGTTTCCCGTTTGGTCCCTCTTCAAGGTTTTTTTTCAGCAGC 488
Qy 490 TGCAGACCCGCTGCCCCACCTATTGCACTGACCTGAGCTGCTGCACTGCACCTTCTACAGC 549
Dy 489 TGCAGACCCGCTGCCCCACCTATTGCACTGAGCTGCTGCACTGCACCTTCTACATC 548
Qy 550 CACACCTGTATGACGCTGAGCCTGCTGCAGGGGCCCTGTTGCACTGAGCCTGCTGTCAGA 609
Dy 549 CACACCTGTATGACGCTGAGCCTGCTGCAGGGGCCCTGTTGCACTGAGCCTGCTGTCAGA 608
Qy 610 GGCACCTGTTGGAGCTGAGCCTGCTGCAGAGGACCTGTTGCACTGAGCCTGCTGTCAGA 669
Dy 609 GGCACCTGTTGGAGCTGAGCCTGCTGCAGAGGACCTGTTGCACTGAGCCTGCTGTCAGA 668
Qy 670 GGCACCTGTTGGAGCTGAGCCTGCTGCAGAGGACCTTCCAGCAGCTGAGCCTGCTGTCAGC 729
Dy 669 GGCACCTGTTGGAGCTGAGCCTGCTGCAGAGGACCTTCCAGCAGCTGAGCCTGCTGTCAGC 728
Qy 730 CAAGCCTGCTGCCCGCAGACCTCACCCCTCTCCCTCTCTTGAACAGGCAAAATCAGTGAAA 789
Dy 729 CAAGCCTGCTGCCCGCAGACCTCACCCCTCTCCCTCTCTTGAACAGGCAAAATCAGTGAAA 788
Qy 790 TTCTCTAGAAGAGTACCATGGGTTCAATTTCTATACGTGACAGAAATAAGTGAATCTAC 849
Dy 789 TTCTCTAGAAGAGTACCATGGGTTCAATTTCTATACGTGACAGAAATAAGTGAATCTAC 848
Qy 850 AAGATTTCTTTCTTTTCCAAAGACTATTTCAATCTGTTGTTATTCAGAGTATTCATCTC 909
Dy 849 AAGATTTCTTTCTTTTCCAAAGACTATTTCAATCTGTTGTTATTCAGAGTATTCATCTC 908
Qy 910 ACTACATTGATTTGTTGTTGTTAGTTT-TTTCCTTGGACTTAATTTATATTTGAAAAACAT 968
Dy 909 ACTACATTGATTTGTTGTTGTTAGTTT-TTTCCTTGGACTTAATTTATATTTGAAAAACAT 968
Qy 969 TGATAATTAATAAATAAATAGATAATTTAGACCAATGG 1008
Dy 969 TGATAATTAATAAATAAATAGATAATTTAGACCAATGG 1008

RESULT 4
AA61352
ID AA61352 standard; cDNA; 959 BP.
XX
AC AA61352;
XX
DT 14-JUL-1999 (first entry)
XX

DE DNA encoding a human secreted protein.
XX
KW Human secreted protein; cancer; tumour; neurodegenerative disorder;
KW developmental abnormality; fetal deficiency; blood disorder; leukemia;
KW immune system disease; autoimmune disease; hepatic disease;
KW renal disease; lymphoma; inflammation; allergy; ischemic shock;
KW Alzheimer's; cognitive disorder; schizophrenia; prostate disease;
KW obesity; osteoclast; osteoporosis; arthritis; malignancy; testes disease;
KW lung disease; thymus disease; digestive disorder; endocrine disorder;
KW infection; AIDS; ss.
XX
OS Homo sapiens.
XX
PN WO9922243-A1.
XX
PD 06-MAY-1999.
XX
XX 23-OCT-1998; 98WO-US22376.
XX
PR 24-OCT-1997; 97US-0063387.
PR 24-OCT-1997; 97US-0062784.
PR 24-OCT-1997; 97US-0063088.
PR 24-OCT-1997; 97US-0063089.
PR 24-OCT-1997; 97US-0063090.
PR 24-OCT-1997; 97US-0063091.
PR 24-OCT-1997; 97US-0063092.
PR 24-OCT-1997; 97US-0063097.
PR 24-OCT-1997; 97US-0063098.
PR 24-OCT-1997; 97US-0063099.
PR 24-OCT-1997; 97US-0063100.
PR 24-OCT-1997; 97US-0063101.
PR 24-OCT-1997; 97US-0063109.
PR 24-OCT-1997; 97US-0063110.
PR 24-OCT-1997; 97US-0063111.
PR 24-OCT-1997; 97US-0063148.
PR 24-OCT-1997; 97US-0063386.
XX
(HUMA-) HUMAN GENOME SCI INC.
XX
PI Brewer LA, Carter KC, Duan DR, Ebner R, Endress CA;
PI Feng P, Florence C, Florence KA, Greene JM, Janat F;
PI Kayw H, Lafleur DW, Moore PA, Ni J, Olsen HS, Rosen CA;
PI Ruben SM, Shi Y, Soppet DR, Wei Y, Young P;
XX
DR WPI; 1999-303069/25.
DR P-PSDB; AAY19472.
XX
PT New isolated human genes and the secreted polypeptides they encode
XX
PS Claim 3; Page 328-329; 546pp; English.
XX
CC The specification describes cDNA sequences (AA61322-X61470) encoding
CC human secreted proteins (AAY19442-Y19590). The polynucleotides and their
CC corresponding secreted polypeptides are useful for preventing, treating
CC or ameliorating medical conditions, e.g. by protein or gene therapy.
CC Pathological conditions can also be diagnosed by determining the amount
CC of the polypeptides in a sample or by determining the presence of
CC mutations in the polynucleotides. Specific uses are described for each
CC of the polynucleotides, based on which tissues they are most highly
CC expressed in, and include developing products for the diagnosis or
CC treatment of cancer, tumours, neurodegenerative disorders, developmental
CC abnormalities and fetal deficiencies, blood disorders, leukemias,
CC diseases of the immune system, autoimmune diseases, hepatic and renal
CC disease, lymphomas, inflammation, allergies, ischemic shock, Alzheimer's
CC and cognitive disorders, schizophrenia, prostate diseases, obesity,
CC disorders involving osteoclasts such as osteoporosis, arthritis or
CC malignancies, diseases of testes, lung or thymus, digestive/endocrine
CC disorders, infections and AIDS. The polypeptides are also useful for
CC identifying their binding partners.
XX
SQ Sequence 959 BP; 240 A; 245 C; 187 G; 286 T; 1 other;
Query Match 88.9%; Score 896.4; DB 20; Length 959;

PT e.g. human blood, kidney, foetal lung, placenta, testes, brain,
PT ovary, pituitary, retina and colon cDNA libraries.
XX
PS Claim 1; Page 284; 618pp; English.
XX
CC The present sequence represents a human expressed sequence tag (EST).
CC The polynucleotide, which is a secreted EST, and the encoded protein
CC are predicted to have useful biological activities which would make
CC them suitable for treating, preventing or ameliorating medical
CC conditions in humans and animals, although no supporting data is
CC given. Suggested activities include nutritional activity, immune
CC stimulating or suppressing activity, haematopoiesis regulating
CC activity, tissue growth activity, activin/inhibin activity,
CC chemotactic/chemokinetic activity, haemostatic and thrombolytic
CC activity, receptor/ligand activity, anti-inflammatory activity,
CC adherin/tumour invasion suppressor activity, tumour inhibition
CC activity. The polynucleotide may also be useful for gene therapy.
XX
SQ Sequence 450 BP; 112 A; 113 C; 78 G; 147 T; 0 other;
Query Match 41.8%; Score 421; DB 20; Length 450;
Best Local Similarity 100.0%; Pred. No. 1.9e-80;
Matches 421; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 49 AAAATAAATTAATAATACATGCGTATTTCTTAAACAATAAATTTATAGTGTAAAT 108
Db |
21 AAAATAAATTAATAATACATGCGTATTTCTTAAACAATAAATTTATAGTGTAAAT 80
QY 109 AATCATAGGTCACAAATCAAGCTTCTCTTTGGCCCTGCATTCATGTGCTTT 168
Db |
81 ATTCATAGGTCACAAATGAAGCTTCTCTTTGGCCCTGCATTCATGTGCTTT 140
QY 169 TCGAAGGAGACAGCGTTCCCTTCATTTGGTGAGGATGACAAATGATGGTCAACCCACT 228
Db |
141 TCGAAGGAGACAGCGTTCCCTTCATTTGGTGAGGATGACAAATGATGGTCAACCCACT 200
QY 229 TCATCATCTCTGAATATTCCTATGGCATACGGAATTTACCACCTCCTCTTATATACG 288
Db |
201 TCATCATCTCTGAATATTCCTATGGCATACGGAATTTACCACCTCCTCTTATATACG 260
QY 289 CCAGTGAATACAGTCCCGAGTTACCTTGGGAATACCTACACTGACACAGGGTTACCTTC 348
Db |
261 CCAGTGAATACAGTCCCGAGTTACCTTGGGAATACCTACACTGACACAGGGTTACCTTC 320
QY 349 GTATCCCTGGATTTCAACTTCTCCTGGATTCCTTATGCTATACATCCGTTGTTTTC 408
Db |
321 GTATCCCTGGATTTCAACTTCTCCTGGATTCCTTATGCTATACATCCGTTGTTTTC 380
QY 409 CTATGCTACTCAGTTCAATGTTCCCTCTCCCTCTAGGGGTTTCCGTTTGTCCCTCC 468
Db |
381 CTATGCTACTCAGTTCAATGTTTCCCTCTCCCTCTAGGGGTTTCCGTTTGTCCCTCC 440
QY 469 T 469
Db |
441 T 441
RESULT 8
AAV89532
ID AAV89532 standard; cDNA; 296 BP.
XX
AC AAV89532;
XX
DT 15-FEB-1999 (first entry)
XX
DE EST clone CP328.
XX
KW Human: secreted protein; expressed sequence tag; EST; haematopoiesis;
KW tissue growth; activin; inhibin; chemotaxis; chemokines; haemostatic;
KW receptor; ligand; thrombolytic; anti-inflammatory; cadherin; anti-tumour;
XX gene therapy; ss.
XX
OS Homo sapiens.

XX W09845436-A2.
PN
XX 15-OCT-1998.
PD
XX
PF 10-APR-1998; 98WO-US069955.
XX
PR 10-APR-1997; 97US-0838821.
XX
PA (GEMY) GENETICS INST INC.
XX
PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D;
PI Racie LA, Spaulding V, Treacy M;
XX WPI; 1999-070077/06.
XX
PT New polynucleotides encoding human secreted proteins - derived from
PT e.g. human blood, kidney, foetal lung, placenta, testes, brain,
PT ovary, pituitary, retina and colon cDNA libraries.
XX
XX Claim 1; Page 247; 618pp; English.
PS
XX The present sequence represents a human expressed sequence tag (EST).
CC The polynucleotide, which is a secreted EST, and the encoded protein
CC are predicted to have useful biological activities which would make
CC them suitable for treating, preventing or ameliorating medical
CC conditions in humans and animals, although no supporting data is
CC given. Suggested activities include nutritional activity, immune
CC stimulating or suppressing activity, haematopoiesis regulating
CC activity, tissue growth activity, activin/inhibin activity,
CC chemotactic/chemokinetic activity, haemostatic and thrombolytic
CC activity, receptor/ligand activity, anti-inflammatory activity,
CC adherin/tumour invasion suppressor activity, tumour inhibition
CC activity. The polynucleotide may also be useful for gene therapy.
XX
SQ Sequence 296 BP; 50 A; 99 C; 82 G; 65 T; 0 other;
Query Match 28.4%; Score 286.4; DB 20; Length 296;
Best Local Similarity 99.7%; Pred. No. 8.7e-52;
Matches 287; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 449 GCTTTCCGTTGTCCCTCCCTTCAAGGTTTTTTTTCAGCAGCTGCAGCACCCGCTGCCCA 508
Db |
1 GGTTCCTCCGTTGTCCCTCCCTTCAAGGTTTTTTTTCAGCAGCTGCAGCACCCGCTGCCCA 60
QY 509 CTAATTCAGCTGAGCTGCTGCAGCTGCACCTCTTACAGCCACACCTGTAGCAGCTGAG 568
Db |
61 CCTATTGAGCTGAGCTGCTGCAGCTGCACCTCTTACAGCCACACCTGTAGCAGCTGAG 120
QY 569 CCTGCTGAGGGGGCCCTGTTGCAGCTGAGCTGCTGCAGAGGCACCTGTGGAGCTGAG 628
Db |
121 CCTGCTGAGGGGGCCCTGTTGCAGCTGAGCTGCTGCAGAGGCACCTGTGGAGCTGAG 180
QY 629 CCTGCTGAGAGGCACCTGTTTGCAGCTGAGCTGCTGCAGAGGCACCTGTGGAGCTGAG 688
Db |
181 CCTGCTGAGAGGCACCTGTTTGCAGCTGAGCTGCTGCAGAGGCACCTGTGGAGCTGAG 240
QY 689 CCAGCTGACAGAGAACCTTTCACCACTGAGCTGCTGCAGAGGCACCTGTGGAGCTGAG 736
Db |
241 CCAGCTGACAGAGAACCTTTCACCACTGAGCTGCTGCAGAGGCACCTGTGGAGCTGAG 288
RESULT 9
AAA30290/c
ID AAA30290 standard; DNA; 3489 BP.
XX
AC AAA30290;
XX
DT 11-SEP-2000 (first entry)
XX
DE Kaposi's sarcoma-associated herpesvirus LANA gene.
XX
KW Kaposi's sarcoma-associated herpesvirus; KSHV; rhadino virus;

KW		latency-associated nuclear antigen; LANA; gamma-2 herpes virus;					
KX		human herpes virus 8; HHV8; rhadino virus cis-acting element; RVCAE;					
KW		Kaposi's sarcoma; primary effusion lymphoma; PEL;					
KX		human immunodeficiency virus; HIV; multicentric Castelman's disease; ds.					
XX							
OS		Kaposi's sarcoma-associated herpesvirus.					
XX							
Key		Location/Qualifiers					
CDS		1..3489					
FT		/*tag= a					
FT		/product= "LANA"					
FT		misc_signal 40..50					
FT		/*tag= b					
FT		/note= "nuclear localisation signal, NLS"					
FT		misc_signal 190..210					
FT		/*tag= c					
FT		/note= "nuclear localisation signal, NLS"					
XX							
PN		WO200029626-A1.					
XX							
XX							
PD		25-MAY-2000.					
XX							
PB		19-NOV-1999; 99WO-US27508.					
XX							
PR		19-NOV-1998; 98US-O109422.					
PR		21-APR-1999; 99US-O298568.					
XX							
PA		(KIEF/) KIEFF E D.					
PA		(BALL/) BALLESTAS M E.					
PA		(KAYE/) KAYE K M.					
XX							
PI		Kieff ED, Ballestas ME, Kaye KM;					
XX							
DR		WPI; 2000-387829/33.					
PS		P-PSDB; AAY96255.					
XX							
PT		Treating or preventing a disease associated with rhodino virus infection in a mammal which includes Kaposi's Sarcoma and Primary Effusion Lymphoma -					
PT							
XX							
PS		Disclosure; Fig 6; 70pp; English.					
XX							
CC		The present sequence is the kaposi's sarcoma-associated herpesvirus, (KSHV) latency-associated nuclear antigen (LANA) gene. KSHV is also known as Human Herpes Virus 8 (HHV8) and belongs to the rhadino virus, or gamma-2 herpes virus class. The LANA protein is necessary for the efficient persistence of rhadino virus DNA in mammalian cells. Persistent rhadino virus infection is implicated in a variety of diseases e.g. Kaposi's Sarcoma (KS), Primary Effusion Lymphoma (PEL) and multicentric Castleman's disease. In addition, KS is a common malignancy in HIV patients. KSHV persists in host cells in a latent form. One of the few genes expressed from the latent viral DNA is LANA. LANA associates with both human chromosomes and with the rhadino virus cis-acting element (RVCAE), thereby providing a tethering function: the KSHV DNA episome is "tied" to the host chromosomes. This allows the viral DNA to persist in the host cell. The present sequence may be used to screen and identify molecules that inhibit LANA interaction with RVCAE, thereby interfering with the latency cycle of this virus. Potential antiviral treatments for the above mentioned diseases may therefore be based on LANA deregulation.					
XX							
SQ		Sequence 3489 BP; 1053 A; 862 C; 1137 G; 437 T; 0 other;					
		Query Match 12.0%; Score 121.2; DB 21; Length 3489;					
		Best Local Similarity 58.7%; Pred. No. 3.5e-16;					
		Matches 210; Conservative 0; Mismatches 148; Indels 0; Gaps 0;					
QY	413	GCTACTCAGTTGAATGTTCCTCCTCTCCCTCTCCTAGGGGTTTCCCCTGCCTCTCA	472				
Db	2188	GCTGCTCATPCCTGCTGCTGCTCTCATCTCGTGCTCATCTCCCTGCTCTGCTCAT	2129				
QY	473	AGTTTTTTTCACAGCTGCACACCCTGCCACCTATTGCAGCTGAGCTGCTGCA	532				
Dd	2128	CTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	2069				

```
XX Sequence 3489 BP; 1053 A; 862 C; 1137 G; 437 T; 0 other;
SQ
  Query Match      12.0%; Score 121.2; DB 22; Length 3489;
  Best Local Similarity 58.7%; Pred. No. 3.5e-16;
  Matches 210; Conservative 0; Mismatches 148; Indels 0; Gaps 0;
  QY 413 GCTACTAGTTGAATGTTCCCTCCTCCTCCTAGGGGTTTCCCGTTTGTCCCTCCTTCA 472
  Db 2188 GCTGCTCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2129
  QY 473 AGGTTTTTTTTCAGCAGCTGCAGACCCCGCTGCCACCCTATTGAGCTGAGCCCTGTGCA 532
  Db 2128 COTGCTGCTGCTGCTCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2069
  QY 533 GCTGCACCTCTTACAGCCACACCTGTAGCAGCTGAGCTGCTGCTGCTGCTGCTGCTGCT 592
  Db 2068 GCTCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2009
  QY 593 GCTGAGCCTGCTGCAGAGGCACCTGTTGGAGCTGAGCCTGCTGCAGAGGCACCTGTGCA 652
  Db 2008 GCTCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1949
  QY 653 GCTGAGCCTGCTGCAGAGGCACCTGTTGGAGTGGAGCCAGCTGCAGAGGAACCTTCACCA 712
  Db 1948 GCTCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1889
  QY 713 GCTGAGCCTGCTACAGCCAAAGCTGCTGCCCGAGAACCTCACCCTTCTCCCTCTCTTG 770
  Db 1888 GCTCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1831
  RESULT 11
  ID ABA93487/c
  AC ABA93487;
  DT 25-APR-2002 (first entry)
  DE Kaposi's sarcoma-associated herpesvirus LANA protein encoding DNA.
  KW Kaposi's sarcoma-associated herpesvirus; KSHV; LANA; RVCAE; PEL;
  KW KSHV terminal repeat; rhadino virus cis acting element; episome;
  KW primary effusion lymphoma; latency-associated nuclear antigen;
  KW gene therapy; gene transfer; gene; ds.
  OS Human herpesvirus 8.
  FH Key Location/Qualifiers
  FT CDS 1..3489
  FT /*tag= a
  FT /product= "LANA protein"
  FT /note= "latency-associated nuclear antigen"
  PN US6322792-B1.
  XX
  XX 27-NOV-2001.
  PF 21-APR-1999; 99US-0298568.
  XX
  PR 19-NOV-1998; 98US-109422P.
  XX
  PA (KIEF/) KIEFF E. D.
  XX
  PI Kieff ED, Ballestas ME, Kaye KM;
  XX
  DR WPI; 2002-153769/20.
  XX
  DR P-PSDB; ABB05621.
  XX
  PT System for episomal retention of plasmids in mammalian cells, useful in
  gene therapy, comprises rhadinoviral LANA and RVCAE sequences -
  XX
```

```
PS Claim 1; Fig 6; 27pp; English.
XX
CC The present invention describes a system (A) for maintaining a plasmid
CC as an episome in mammalian cells, comprising the rhadinoviral sequence
CC LANA (latency-associated nuclear antigen) of 3489 base pairs (see
CC ABA93487, S1) expressed in the cell, and the rhadinoviral sequence RVCAE
CC (rhadinoviral cis-acting element) of 801 base pairs (see ABA93488, S2)
CC present in the plasmid. Also describes is a method for maintaining a
CC closed circular DNA in a cell by expressing (S1) in the cells and having
CC (S2) as a cis-acting and maintenance sequence in the DNA. (A) is
CC particularly used in gene therapy (or other gene transfer applications)
CC that uses mammalian cells in which LANA is expressed. (A) improves
CC persistence of gene therapy vectors in cells. The present sequence
CC encodes Kaposi's sarcoma-associated herpesvirus (KSHV, also called human
CC herpesvirus 8) LANA protein, which is used in the exemplification of the
CC present invention.
XX
SQ Sequence 3489 BP; 1053 A; 862 C; 1137 G; 437 T; 0 other;
  Query Match      12.0%; Score 121.2; DB 24; Length 3489;
  Best Local Similarity 58.7%; Pred. No. 3.5e-16;
  Matches 210; Conservative 0; Mismatches 148; Indels 0; Gaps 0;
  QY 413 GCTACTAGTTGAATGTTCCCTCCTCCTCCTAGGGGTTTCCCGTTTGTCCCTCCTTCA 472
  Db 2188 GCTGCTCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2129
  QY 473 AGGTTTTTTTTCAGCAGCTGCAGACCCCGCTGCCACCCTATTGAGCTGAGCCTGTGCA 532
  Db 2128 COTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2069
  QY 533 GCTGCACCTCTTACAGCCACACCTGTAGCAGCTGAGCCTGCTGCAGAGGCACCTGTGCA 592
  Db 2068 GCTCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2009
  QY 593 GCTGAGCCTGCTGCAGAGGCACCTGTTGGAGCTGAGCCTGCTGCAGAGGCACCTGTGCA 652
  Db 2008 GCTCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1949
  QY 653 GCTGAGCCTGCTGCAGAGGCACCTGTTGGAGTGGAGCCAGCTGCAGAGGAACCTTCACCA 712
  Db 1948 GCTCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1889
  QY 713 GCTGAGCCTGCTACAGCCAAAGCTGCTGCCCGAGAACCTCACCCTTCTCCCTCTCTTG 770
  Db 1888 GCTCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1831
  RESULT 12
  AAV73805
  ID AAV73805 standard; DNA; 32207 BP.
  XX
  XX AAV73805;
  AC AAV73805;
  DT 25-FEB-1999 (first entry)
  XX
  XX KSHV LUR DNA (nucleotides 105,301-137,507).
  DE
  DE
  DE
  KW Kaposi's sarcoma; acquired immune deficiency syndrome; AIDS; DHFR; Bcl-2;
  KW dihydrofolate reductase; LUR; long unique region; vaccine; prophylaxis;
  KW diagnosis; treatment; HHV8; capsid protein IV; tegument protein IV;
  KW glycoprotein; kaposin; cyclin D; immediate early protein; IEP; OX-2;
  KW v-adh; G-protein coupled receptor; FGARAT; ds.
  XX
  OS Kaposi's sarcoma-associated herpesvirus.
  XX
  PN US5849564-A.
  XX
  PD 15-DEC-1998.
  XX
  PF 29-NOV-1996; 96US-0770379.
  XX
  PR 29-NOV-1996; 96US-0770379.
```

```
XX (UYCO ) UNIV COLUMBIA NEW YORK.
XX PA
XX Bohenzky RA, Chang Y, Edelman IS, Moore PS, Russo JJ;
XX WPI; 1999-069741/06.
XX Kaposi's sarcoma-associated herpes virus nucleic acid - encodes
PT di-hydro:folate reductase and is useful for treatment, prophylaxis
PT or diagnosis of Kaposi's sarcoma
XX
XX PS Disclosure: Column 155-182; 109pp; English.
XX
XX This sequence is a fragment of the Kaposi's sarcoma-associated
CC herpesvirus (KSHV) LUR (long unique region). This fragment contains
CC coding regions for ORF65 which encodes capsid protein IV, ORF66, ORF67
CC which encodes tegument protein IV, ORF68 which encodes a glycoprotein,
CC ORF69, K12 which encodes Kapsin, K13, ORF72 which encodes cyclin D,
CC ORF73 which encodes immediate early protein (IEP), K14 which encodes
CC OX-2 (V-adh), ORF74 which encodes G-protein coupled receptor, ORF75
CC which encodes tegument protein/FGFARAT, K15. KSHV is a new human
CC Herpesvirus (HHV8) believed to cause Kaposi's sarcoma (KS) which is the
CC most common form of neoplasia occurring in persons with acquired immune
CC deficiency syndrome (AIDS). The DHFR protein is useful for vaccination,
CC prophylaxis, diagnosis and treatment of a subject with Kaposi's sarcoma
CC and for detecting expression of a DNA virus associated with Kaposi's
CC sarcoma in a cell.
XX
XX SQ Sequence 32207 BP; 7229 A; 9156 C; 8713 G; 7109 T; 0 other;
Query Match 12.0%; Score 121.2; DB 20; Length 32207;
Best Local Similarity 58.7%; Pred. No. 7.6e-16;
Matches 210; Conservative 0; Mismatches 148; Indels 0; Gaps 0;
QY 413 GCTACTCAGTGAATGTTCCCTCCTCCTAGGGGTTTCCCGTTTGCCCTCCTTCA 472
Db 19809 GGTGCTCATCTGCTGCTGCTGCTCATCTGCTGCTGCTGCTGCTGCTGCTCAT 19868
QY 473 AGGTTTTTTTTCAGCAGCTGCAGCACCCGCTGCCACACCTATTGAGCTGAGCGCTGCA 532
Db 19869 CTTGCTGCTCCTGCTCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 19928
QY 533 GGTGACCTCTTTACAGCCACACCTGTAGCAGCTGAGCTGCTGAGGGGCCCTGTTGCA 592
Db 19929 GGTATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 19988
QY 593 GGTGAGCTGCTGAGAGGACCTGTTGGAGCTGAGCTGAGCTGCTGAGAGGACCTGTTGCA 652
Db 19989 GGTATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 20048
QY 653 GGTGAGCTGCTGAGAGGACCTGTTGGAGTGGAGCCAGCTGAGAGGACCTTCACCA 712
Db 20049 GGTATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 20108
QY 713 GGTGAGCTGCTTACAGCAAGCTGCTGCCCGAGAACCTCACCTCTTCCTCTCTTG 770
Db 20109 GGTATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 20166
RESULT 13
AAV19941
ID AAV19941 standard; DNA; 137507 BP.
XX
XX AC AAV19941;
XX
XX DT 03-AUG-1998 (first entry)
XX
XX DE KSHV long unique coding region and terminal repeat.
XX
XX KW KSHV; HHV8; human herpes virus 8; macrophage inflammatory protein II;
KW interleukin-6; IL-6; interferon regulatory factor; rheumatoid arthritis;
KW complement-binding protein; glycoprotein; capsid protein IV; infection;
KW immediate early protein; Kaposi's sarcoma; protective vaccine; lymphoma;
```

```
KW lymphoproliferative disease; leukaemia; splenomegaly; mycosis fungoides;
KW HIV immune status; anti-inflammatory agent; therapy; ds.
OS Kaposi's sarcoma-associated herpes virus.
XX
XX FH Location/Qualifiers
XX CDS 1142..2794
XX FT /*tag= a
XX FT /product= complement-binding protein
XX CDS 8699..11236
XX FT /*tag= b
XX FT /product= glycoprotein B
XX CDS complement (17261..17875)
XX FT /*tag= c
XX FT /product= interleukin 6
XX CDS complement (21548..21832)
XX FT /*tag= d
XX FT /product= macrophage inflammatory protein II
XX CDS complement (27137..27424)
XX FT /*tag= e
XX FT /product= interferon regulatory factor 1
XX CDS 28661..29741
XX FT /*tag= f
XX FT /product= protein Tl.1
XX CDS complement (58976..60175)
XX FT /*tag= g
XX FT /product= glycoprotein M
XX CDS complement (69412..69915)
XX FT /*tag= h
XX FT /product= glycoprotein L
XX CDS complement (88410..88910)
XX FT /*tag= i
XX FT /product= interferon regulatory factor 2
XX CDS 89600..90541
XX FT /*tag= j
XX FT /product= interferon regulatory factor 3
XX CDS 90173..90643
XX FT /*tag= k
XX FT /product= glycoprotein x
XX CDS complement (93636..94127)
XX FT /*tag= l
XX FT /product= interferon regulatory factor 4
XX CDS complement (111931..112443)
XX FT /*tag= m
XX FT /product= capsid protein IV
XX CDS complement (123808..127296)
XX FT /*tag= n
XX FT /product= immediate early protein
XX
XX WO9804576-A1.
XX
XX PD 05-FEB-1998.
XX
XX PF 22-JUL-1997; 97WO-US13346.
XX
XX PR 29-NOV-1996; 96US-0757669.
XX PR 25-JUL-1996; 96US-0686243.
XX PR 25-JUL-1996; 96US-0686349.
XX PR 25-JUL-1996; 96US-0686350.
XX PR 25-JUL-1996; 96US-0687253.
XX PR 25-JUL-1996; 96US-0688814.
XX PR 05-SEP-1996; 96US-0708678.
XX PR 10-OCT-1996; 96US-0728323.
XX PR 13-NOV-1996; 96US-0747887.
XX PR 13-NOV-1996; 96US-0748640.
XX
XX (UYCO ) UNIV COLUMBIA NEW YORK.
XX
XX PI Bohenzky RA, Chang Y, Edelman IS, Moore PS, Russo JJ;
XX WPI; 1998-130615/12.
XX
XX PR New nucleic acid encoding Kaposi's sarcoma associated herpes virus
XX
```



```
PT proteins - useful for, e.g. detecting levels of HHV8 in, and
PT preparation of vaccines for treatment of, HIV patients
XX
PS Example 2; Page 135-203; 230pp; English.
XX
XX This sequence represents the long unique region and terminal repeat of
CC the Kaposi's sarcoma-associated herpes virus (KSHV). KSHV is also known
CC as human herpes virus 8 (HHV8). This sequence contains the DNAs of the
CC invention which encode KSHV polypeptides selected from: (a) viral
CC macrophage inflammatory protein (MIP) II; (b) viral interleukin-6 (IL-6);
CC (c) viral IRF 1; (d) complement-binding protein; glycoproteins B, M or L;
CC (d) capsid protein IV encoded by ORF65; and (e) immediate early protein
CC encoded by ORF73. Labelled probes for the nucleic acid, proteins encoded
CC by it, and antibodies (Ab) specific for the proteins are useful for
CC detecting HHV8, specifically for diagnosis of Kaposi's sarcoma, in body
CC fluids or tissue samples. HHV8 infections can be treated with antisense
CC or triplex forming molecules or agents that bind specifically to the
CC protein. Ab may be used for prophylaxis or treatment of HHV8 infection,
CC while the protein can be used in protective vaccines. Ab may also be used
CC to differentiate between lymphomas, and HHV8 may be implicated in many
CC other lymphoproliferative diseases such as lymphomas, leukaemia,
CC splenomegaly and mycosis fungoides. Cells and animals containing the
CC nucleic acid are useful for drug screening. HHV8-derived peptides can be
CC used as targets for antiviral drugs, e.g. dihydrofolate reductase gene
CC can be inhibited with methotrexate. These can also be used to determine
CC the immune status of a patient infected with HIV. HHV8 derived protein
CC viral MIP III may be used as an anti-inflammatory agent for,
CC e.g. treating rheumatoid arthritis. This sequence is stated as containing
CC 81 open reading frames.
XX
SQ Sequence 137507 BP; 32579 A; 37795 C; 35758 G; 31375 T; 0 other;

Query Match 12.0%; Score 121.2; DB 19; Length 137507;
Best Local Similarity 58.7%; Pred. No. 1.2e-15;
Matches 210; Conservative 0; Mismatches 148; Indels 0; Gaps 0;

QY 413 GCTACTCAGTTGAATGTTCCTCTCCTCCTCCTAGGGGTTTCCCGTTTGTCCCTCCTTCA 472
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 125109 GCTGCTCATCCTGCTGCTGCTGCTCATCCTGCTGCTGCTCATCCTGCTGCTCAT 125168

QY 473 AGGTTTTCAGCAGCTGACGACCGCGTGCCTCCACCTATTGACGCTGACCTGCTGCA 532
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 125169 CCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 125228

QY 533 GCTGCACCTCTTACAGCACACCTGTAGCAGCTGAGCTGCTGAGGGGCCCTGTTTGA 592
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 125229 GCTCATCTGCTGCTGCTGCTCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 125288

QY 593 GCTGAGCTGCTGCAGAGGACACCTGTTGGAGCTGAGCTGCTGCGAGAGCACCTGTTGCA 652
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 125289 GCTCATCTGCTGCTGCTGCTCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 125348

QY 653 GCTGAGCTGCTGCAGAGGACACCTGTTGGAGTGAGGACGCTGCAGAGAACCTTCACCA 712
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 125349 GCTCATCTGCTGCTGCTGCTCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 125408

QY 713 GCTGAGCTGCTACAGCCAGAGCTGCTGCCAGAACCTCAGCCCTTCTCCCTCTCTTGG 770
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 125409 GCTCATCTGCTGCTGCTCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 125466

RESULT 14
AAC84677/c
ID AAC84677 standard; DNA; 5120 BP.
XX
AC AAC84677;
XX
XX 20-APR-2001 (first entry)
XX
XX DNA sequence of human PPP2R2B.
DE
XX Microsatellite marker; human; chromosome 5q31-33; D5S402; WI-6763;
KW CAG trinucleotide repeat; spinocerebellar ataxia 12; SCA12; PPP2R2Bbeta;
KW
```

```
KW brain; protein phosphatase PP2A; ds.
XX
XX Homo sapiens.
XX
FH Key Location/Qualifiers
FT repeat_region 2088..2366
FT /tag= a
FT /rpt_type= "CAG"
FT mRNA 2473..3093
FT /tag= b
FT 5'UTR 2473..3023
FT /tag= c
FT CDS 3024..3093
FT /tag= d
FT /product= "PPP2R2B"
FT /note= "protein_id= AAF74024.1"
XX
XX WO200078943-A2.
PN
XX 28-DEC-2000.
PD
XX 16-JUN-2000; 2000WO-US40213.
XX
XX 18-JUN-1999; 990US-0140176.
XX
XX (UYJO ) UNIV JOHNS HOPKINS.
XX
XX Margolis RL, Ross CA, Holmes SE;
PI
XX WPI; 2001-102718/11.
DR P-PSDB; AAB48340.
XX
XX New polynucleotide for use in diagnosing spinocerebellar ataxia 12,
PT comprises a microsatellite marker having a variable number of CAG
PT trinucleotide repeats located on human chromosome 5q31-33
XX
XX Disclosure; Page 21-23; 23pp; English.
PS
XX The invention relates to a microsatellite marker located on human
XX chromosome 5q31-33 between markers D5S402 and WI-6763, comprising a
XX variable number of CAG trinucleotide repeats. The marker is useful for
XX diagnosing spinocerebellar ataxia 12 (SCA12) by determining the number of
XX CAG trinucleotide repeats in a 5'-untranslated region of an allele of
XX PPP2R2Bbeta, a brain specific regulatory subunit of protein phosphatase
XX PP2A, where if a number greater than 40 is determined, an expansion which
XX is associated with SCA12 is indicated. The presence of expanded
XX trinucleotide repeats on chromosome 5q31-33 is useful for predicting or
XX diagnosing SCA12. Also provided are primers useful for amplifying the
XX microsatellite marker associated with SCA12 and for diagnosing SCA12 and
XX a probe useful for detecting expansions in a microsatellite marker
XX located on human chromosome 5q31-33. The present sequence represents
XX a PPP2R2B DNA sequence.
SQ Sequence 5120 BP; 1336 A; 1267 C; 1268 G; 1249 T; 0 other;

Query Match 10.5%; Score 106; DB 22; Length 5120;
Best Local Similarity 61.0%; Pred. No. 7.1e-13;
Matches 172; Conservative 0; Mismatches 110; Indels 0; Gaps 0;

QY 485 GCAGCTGCAGACACCGCTGCCCGACCTATTGACGTGAGCTGCTGCAGCTGCACCTCTT 544
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2370 GCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2311

QY 545 ACAGCCACACCTGTAGCAGCTGAGCTGCTGAGGGGCCCTGTTGCAGCTGAGCTGCT 604
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2310 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2251

QY 605 GCAGAGGCACCTGTTGGAGCTGAGCTGCTGAGAGGACACCTGTTGCAGCTGAGCTGCT 664
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2250 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2191

QY 665 GCAGAGGCACCTGTTGGAGTGAGGACGCTGCAGAGGAACCTTCACCACTGAGCTGCT 724
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```



```

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/728.323A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 0575/52268/JPW/MSC/SKS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-278-0400
; TELEFAX: 212-391-0525
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3489 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..3489
; US-08-728-323A-1

Query Match 12.0%; Score 121.2; DB 2; Length 3489;
Best Local Similarity 58.7%; Pred. No. 5.1e-20;
Matches 210; Conservative 0; Mismatches 148; Indels 0; Gaps 0;

QY 413 GCTACTCAGTTGAATGTTCCCTCTCCCTCTAGGGGTTTCCCGTTTGTCCCTCCTTCA 472
||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 2188 GCTGCTCATCCTGCTGCTGCTCATCTGCTGCTCATCTGCTGCTGCTGCTGCTCAT 2129
QY 473 AGGTTTTTTTACGAGCTGCAGACCCGCTGCCCCACCTATTGACGCTGAGCGTGTGCA 532
||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 2128 CTGCTGCTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2069
QY 533 GCTGCACCTCTTACAGCCACACCTGTAGCAGCTGAGCGTGTGAGGGGCGCCCTGTTGCA 592
||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 2068 GCTCATCTCTGCTGCTGCTCATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2009
QY 593 GCTGAGCCTGCTGAGAGGACCTGTTGGAGCTGAGCCTGCTGCGAGGACCTGTTGCA 652
||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 2008 GCTCATCTCTGCTGCTGCTCATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1949
QY 653 GCTGAGCCTGCTGCAGAGGCACTGTTGGAGTGGAGCGACCTGCAGAGGAACCTCACCA 712
||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 1948 GCTCATCTCTGCTGCTGCTCATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1889
QY 713 GCTGAGCCTGCTACAGCAAGCCTGCTGCGCCAGAACCTCACCTTCTCCCTCTCTTG 770
||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 1888 GCTCATCTCTGCTGCTGCTCATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1831

RESULT 4
US-09-298-568-1/C
; Sequence 1, Application US/09298568
; Patent No. 6322792
; GENERAL INFORMATION:
; APPLICANT: Kieff, Elliott D.
; APPLICANT: Ballestas, Mary E.
; APPLICANT: Kaye, Kenneth M.
; TITLE OF INVENTION: RHADNO VIRUS LANA ACTS IN TRANS ON A UNIT OF RHADNO
; TITLE OF INVENTION: VIRUS DNA TO MEDIATE EFFICIENT EPISOME PERSISTENCE
; FILE REFERENCE: 16412-10001R
; CURRENT APPLICATION NUMBER: US/09/298, 568
; CURRENT FILING DATE: 1999-04-21
; EARLIER APPLICATION NUMBER: US 60/109,422
; EARLIER FILING DATE: 1998-11-19
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 3489

```

```
;
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-770-379-20

Query Match 12.0%; Score 121.2; DB 2; Length 32207;
Best Local Similarity 58.7%; Pred. No. 1.5e-19;
Matches 210; Conservative 0; Mismatches 148; Indels 0; Gaps 0;

Qy 413 GCTACTCAGTTGAATGTTCCCTCTCTCCTAGAGGTTTCCCGTTTGTCCCTCTTCA 472
||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 19809 GCTGCTATCCCTGCTGCTGCTCATCTCTGCTGCTCATCTCTGCTGCTGCTCAT 19868

Qy 473 AGGTTTTTTCAGCAGCTGCAGCACCCGCTGCCACCTATTGCAGCTGAGCTGTGCA 532
||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 19869 CTTGCTCTCTGCTCATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 19928

Qy 533 GCTGCACCTCTTACAGCCACACCTGTAGCAGCTGAGCCTGCTGCAGGGGCCCTTTGCA 592
||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 19929 GCTCATCTGCTGCTGCTCATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 19988

Qy 593 GCTGAGCCTGCTGCAGAGCACCTGTTGGAGCTGAGCCTGCTGCAGAGGACCTGTTGCA 652
||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 19989 GCTCATCTGCTGCTGCTCATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 20048

Qy 653 GCTGAGCCTGCTGCAGAGCACCTGTTGGAGTGGAGGAGCAGCTGCAGAGGAACCTTCACCA 712
||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 20049 GCTCATCTGCTGCTGCTCATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 20108

Qy 713 GCTGAGCCTGCTACAGCCAAAGCCTGCTGCCCCAGAACCTCACCCCTTCTCCCTCTTG 770
||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 20109 GCTCATCTGCTGCTGCTCATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 20166

RESULT 6
US-08-757-669A-20
; Sequence 20, Application US/08757669A
; Patent No. 6183751
; GENERAL INFORMATION:
; APPLICANT: Chang, Yuan
; APPLICANT: Bohenzky, Roy A.
; APPLICANT: Russo, James J.
; APPLICANT: Edelman, Isidore S.
; APPLICANT: Moore, Patrick S.
; TITLE OF INVENTION: UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS
; TITLE OF INVENTION: SEQUENCES AND USES THEREOF
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESS: Cooper & Dunham LLP
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/757,669A
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 45185-F
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 32207 base pairs
```

```
;
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-757-669A-20

Query Match 12.0%; Score 121.2; DB 4; Length 32207;
Best Local Similarity 58.7%; Pred. No. 1.5e-19;
Matches 210; Conservative 0; Mismatches 148; Indels 0; Gaps 0;

Qy 413 GCTACTCAGTTGAATGTTCCCTCTCTCCTAGAGGTTTCCCGTTTGTCCCTCTTCA 472
||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 19809 GCTGCTATCCCTGCTGCTGCTCATCTCTGCTGCTGCTCATCTCTGCTGCTGCTCAT 19868

Qy 473 AGGTTTTTTCAGCAGCTGCAGCACCCGCTGCCACCTATTGCAGCTGAGCTGTGCA 532
||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 19869 CTTGCTCTCTGCTCATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 19928

Qy 533 GCTGCACCTCTTACAGCCACACCTGTAGCAGCTGAGCCTGCTGCAGGGGCCCTTTGCA 592
||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 19929 GCTCATCTGCTGCTGCTCATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 19988

Qy 593 GCTGAGCCTGCTGCAGAGCACCTGTTGGAGCTGAGCCTGCTGCAGAGGACCTGTTGCA 652
||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 19989 GCTCATCTGCTGCTGCTCATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 20048

Qy 653 GCTGAGCCTGCTGCAGAGCACCTGTTGGAGTGGAGGAGCAGCTGCAGAGGAACCTTCACCA 712
||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 20049 GCTCATCTGCTGCTGCTCATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 20108

Qy 713 GCTGAGCCTGCTACAGCCAAAGCCTGCTGCCCCAGAACCTCACCCCTTCTCCCTCTTG 770
||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 20109 GCTCATCTGCTGCTGCTCATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 20166

RESULT 7
US-09-230-371A-20
; Sequence 20, Application US/09230371A
; Patent No. 6348586
; GENERAL INFORMATION:
; APPLICANT: Chang, Yuan
; APPLICANT: Bohenzky, Roy A.
; APPLICANT: Russo, James J.
; APPLICANT: Edelman, Isidore S.
; APPLICANT: Moore, Patrick S.
; TITLE OF INVENTION: UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS SEQUENCES AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: 45185-G-PCT-US
; CURRENT APPLICATION NUMBER: US/09/230,371A
; CURRENT FILING DATE: 1999-11-17
; PRIOR APPLICATION NUMBER: PCT/US97/13346
; PRIOR FILING DATE: 1997-07-22
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 20
; LENGTH: 32207
; TYPE: DNA
; ORGANISM: Kaposi's sarcoma-associated herpesvirus
US-09-230-371A-20

Query Match 12.0%; Score 121.2; DB 4; Length 32207;
Best Local Similarity 58.7%; Pred. No. 1.5e-19;
Matches 210; Conservative 0; Mismatches 148; Indels 0; Gaps 0;

Qy 413 GCTACTCAGTTGAATGTTCCCTCTCTCCTAGAGGTTTCCCGTTTGTCCCTCTTCA 472
||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 19809 GCTGCTATCCCTGCTGCTGCTCATCTCTGCTGCTGCTCATCTCTGCTGCTGCTCAT 19868

Qy 473 AGGTTTTTTCAGCAGCTGCAGCACCCGCTGCCACCTATTGCAGCTGAGCTGTGCA 532
||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 19869 CTTGCTCTCTGCTCATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 19928

Qy 533 GCTGCACCTCTTACAGCCACACCTGTAGCAGCTGAGCCTGCTGCAGGGGCCCTTTGCA 592
```



```

; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 234 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
US-08-469-802B-3

Query Match      8.7%; Score 87.6; DB 1; Length 234;
Best Local Similarity 63.1%; Pred. No. 1.9e-12;
Matches 135; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

QY 482 TCAGCAGCTGCAGCACCCTGCCCCACCTATTGACGCTGAGCCTGCTGCAGCTGCACCT 541
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 214 TGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 155

QY 542 CTTACAGCCACACCTGTAGCAGCTGAGCCTGCTGCAGGAGGCCCCCTGTTGCAGCTGAGCCT 601
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 154 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 95

QY 602 GCTGCAGAGGACACCTGTTGGAGCTGAGCCTGCTGCAGGAGGCCCCCTGTTGCAGCTGAGCCT 661
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 94 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 35

QY 662 GCTGCAGAGGACACCTGTTGGAGTGGAGCGAGCAGCTG 695
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 34 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1

RESULT 11
US-08-267-803B-3/c
; Sequence 3, Application US/08267803B
; Patent No. 5834183
; GENERAL INFORMATION:
; APPLICANT: Orr, Harry T.
; APPLICANT: Ranum, Laura P.W.
; APPLICANT: Chung, Ming-yi
; APPLICANT: Zoghbi, Huda Y.
; TITLE OF INVENTION: Gene Sequence for Spinocerebellar Ataxia
; Patent No. 5834183
; TITLE OF INVENTION: Type 1 and Method for Diagnosis
; NUMBER OF SEQUENCES: 85
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Mueiting, Raasch, Gebhardt & Schwappach, P.A.
; STREET: P.O. Box 581415
; CITY: Minneapolis
; STATE: MN
; COUNTRY: USA
; ZIP: 55458-1415
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/267,803B
; FILING DATE: 28-JUN-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: McCormack, Myra H.
; REGISTRATION NUMBER: 36,602
; REFERENCE/DOCKET NUMBER: 110.00030120
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 612-305-1217
; TELEFAX: 612-305-1228
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 234 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA

```

```

US-08-267-803B-3

Query Match      8.7%; Score 87.6; DB 2; Length 234;
Best Local Similarity 63.1%; Pred. No. 1.9e-12;
Matches 135; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

QY 482 TCAGCAGCTGCAGCACCCTGCCCCACCTATTGACGCTGAGCCTGCTGCAGCTGCACCT 541
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 214 TGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 155

QY 542 CTTACAGCCACACCTGTAGCAGCTGAGCCTGCTGCAGGAGGCCCCCTGTTGCAGCTGAGCCT 601
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 154 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 95

QY 602 GCTGCAGAGGACACCTGTTGGAGCTGAGCCTGCTGCAGGAGGCCCCCTGTTGCAGCTGAGCCT 661
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 94 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 35

QY 662 GCTGCAGAGGACACCTGTTGGAGTGGAGCGAGCAGCTG 695
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 34 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1

RESULT 12
US-08-403-379A-2
; Sequence 2, Application US/08403379A
; Patent No. 5756662
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DETECTION
; TITLE OF INVENTION: OF T. CRUZI INFECTION
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/403,379A
; FILING DATE: 14-MAR-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.406
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 786 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
US-08-403-379A-2

Query Match      8.6%; Score 86.6; DB 1; Length 786;
Best Local Similarity 56.0%; Pred. No. 6.1e-12;
Matches 164; Conservative 0; Mismatches 129; Indels 0; Gaps 0;

QY 485 GCAGCTGCAGCACCCTGCTGCCACCTATTGACGCTGAGCCTGCTGCAGCTGCACCTTTT 544
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 433 GCCTGCTGCACCTGCGAAGGCGCTGCTGCACCTGCGAAGGCTGCTGCTGCACCTGCG 492

QY 545 ACAGCCACACCTGTAGCAGCTGAGCCTGCTGCAGGAGGCCCCCTGTTGCAGCTGAGCCTGCT 604
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```


Db 493 AAGGCGCTGCTGCACCTGCGAAGGCTGCTGTCGACCTGCGAAGGCTGCTACTGCACCT 552
QY 605 GCAGAGGACCTGTTGGAGCTGAGCCTGCTGCAGAGGACACCTGTTGCACTGAGCCTGCT 664
Db 553 GCGAAGGCTGCTGCTGCACCTGCCAAGACCGTGTGTCACCTGCGAAGGCTGCTGCACCT 612
QY 665 GCAGAGGACCTGTTGGAGTGAGCGAGCTGCAGAGGAACTTACCAAGCTGAGCCTGCT 724
Db 613 GCGAAGGCGCTGCTGCACCTGCGAAGGCGCTACTGCACCTGCGAAGGCTGCTGCTGCA 672
QY 725 ACAGCAAGCCTGCTGCCCGCAAGACCTCACCTTCTCCCTCTCTTTGAACAGGC 777
Db 673 CCTGCGAAGGCGCTACTGCACTGCGAAGGCTGCTACTGCACCTGCGAAGGC 725

RESULT 13
US-08-929-414-2
; Sequence 2, Application US/08929414
; Patent No. 5942403
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Haughton, Raymond
; APPLICANT: Sheiky, Yasir A.W.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DETECTION
; OF T. CRUZI INFECTION
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/929,414
; FILING DATE: 15-SEP-1997
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.406C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 786 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-929-414-2

Query Match 8.6%; Score 86.6; DB 2; Length 786;
Best Local Similarity 56.0%; Pred. No. 6.1e-12;
Matches 164; Conservative 0; Mismatches 129; Indels 0; Gaps 0;
QY 485 GCAGCTGCAGCACCGCTGCCCACTATTGCAGCTGAGCCTGCTGCAGCTGCACTGCTT 544
Db 433 GCCGCTGCTGCACCTGCGAAGGCGCTGCTGCACCTGCGAAGGCTGCTGCTGCACCTGCG 492
QY 545 ACAGCCACACCTGTAGCACTGAGCCTGCTGCAGGGGCCCCCTGTGTGCACTGAGCCTGCT 604
Db 493 AAGGCGGCTGCTGCACCTGCGAAGGCTGCTGCTGCACCTGCGAAGGCTGCTACTGCACCT 552
QY 605 GCAGAGGACCTGTTGGAGCTGAGCCTGCTGCAGAGGACACCTGTTGCACTGAGCCTGCT 664
Db 553 GCGAAGGCTGCTGCTGCACCTGCCAAGACCGCTGCTGCACCTGCGAAGGCTGCTGCACCT 612

QY 665 GCAGAGGACCTGTTGGAGTGAGCGAGCTGCAGAGAACCTTACCAGCTGAGCCTGCT 724
Db 613 GCGAAGGCGCTGCTGCACCTGCGAAGGCGCTACTGCACCTGCGAAGGCTGCTGCTGCA 672
QY 725 ACAGCCACACCTGCTGCCCGCAAGACCTCACCTTCTCCCTTCTTGAACAGGC 777
Db 673 CCTGCGAAGGCGCTACTGCACCTGCGAAGGCTGCTACTGCACCTGCGAAGGC 725
RESULT 14
US-09-249-585A-2/c
; Sequence 2, Application US/09249585A
; Patent No. 6417002
; GENERAL INFORMATION:
; APPLICANT: Horlick, Robert
; TITLE OF INVENTION: METHOD FOR MAINTENANCE AND SELECTION OF EPISOMES
; FILE REFERENCE: 0867/0D905
; CURRENT APPLICATION NUMBER: US/09/249,585A
; CURRENT FILING DATE: 1999-02-11
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 1926
; TYPE: DNA
; ORGANISM: Epstein Barr Virus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1926)
; OTHER INFORMATION: coding strand of EBNA-1 DNA
US-09-249-585A-2

Query Match 8.5%; Score 85.2; DB 4; Length 1926;
Best Local Similarity 51.9%; Pred. No. 2e-11;
Matches 192; Conservative 0; Mismatches 178; Indels 0; Gaps 0;
QY 397 CCGTGGTTCCTCCCTTACTACTACTCATGTTGAATGTTCTCTCTCTCCCTCCTAGGGTTTCC 456
Db 963 CCCTCTGCTCCTGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 904
QY 457 GTTGTGCTCCCTTCAAGGTTTTCAGCAGCTGCAGACACCGCTGCCCACTATTGC 516
Db 903 TCCTGCCCTCTCTGCTCTCTGCCCTCTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 844
QY 517 AGCTGAGCCTGCTGCAGCTGCACCTCTTACAGCCACACCTGTAGCAGCTGAGCCTGCTGC 576
Db 843 TCCTGCTCTCTGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 784
QY 577 AGGGGCCCCCTGTTGCACTGAGCCTGCTGCAGAGGACACCTGTTGGAGCTGAGCCTGCTGC 636
Db 783 TGCCCCCTCTGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 724
QY 637 AGAGGACCTGTTGCAGCTGAGCCTGCTGCAGAGGACACCTGTTGGAGTGGAGGACGCTGC 696
Db 723 TCCTGCCCTCTCTGCTCTCTGCCCTCTCTGCTCTCTGCCCTCTCTCTCTCTCTCTCTCTCTCT 664
QY 697 AGAGAACCTTACCAGCTGAGCCTGCTACAGCCCAAGCCTGCTGCCCGCCAGAACCTCACCC 756
Db 663 TCCTGCTCTCTGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 604
QY 757 TTCTCCCTCT 766
Db 603 CCCTCTCTCT 594
RESULT 15
US-09-050-863-2/c
; Sequence 2, Application US/09050863
; Patent No. 611411
; GENERAL INFORMATION:
; APPLICANT: Lao, Ying
; APPLICANT: Hiang, Betty
; APPLICANT: Pavan, Don
; TITLE OF INVENTION: Mammalian Protein Interaction Cloning

[illegible]

Db 121 C N T A Y G G N A T H M G N A A Y T T N C C C N C C N T N T A Y T A Y M G N C C N G T N A A Y A C N G I N C C N 180
Qy 308 A G T T A C C C T G G G A A T A C T A C A G A C A C A G G G T T A C C T T C G T A T C C C T G A T T C T A A C T 367
Db 181 W S N T A Y C N G N A A Y A C N T A Y A C N G A Y A C N G N Y T N C N S N A Y C C N T G A T T N A C N 240
Qy 368 T C T C C T G A T T C C C C T A T G T A T A C A T C C G T G G T T T T C C C T T A G C T A C T A C A T T G A A T 427
Db 241 W S N C N G N T Y C C N T A Y G T N T A Y C A T A T H M G N G N T Y C C N Y T N G C N A C N A R Y T N A A Y 300
Qy 428 G T T C C T C C T C C C T A G G G T T T C C C G T T T G C C T C C T C A A G G T T T T T T C A G C A 487
Db 301 G T N C C N C N T N C C N M G N G T T Y C C N T Y G T N C C N C N S N M G N T T Y T T W S N G C N 360
Qy 488 G C T C A G A C C C G T G C C C A C C A T T A T G C A G C T G A G C C T G C T G C A G C T G C A C C T C T T A C A 547
Db 361 G C N C N G C N C N G C N C C N C C N A T H G C N G C N G A R C N C N G C N G C N G C N C C N Y T N A C N 420
Qy 548 G C C A C A C C T G T A G C A G C T G A G C T G C T G C A G G G C C C C T T T G C A G C T G A G C C T G C T G C A 607
Db 421 G C N A C N C N G T N G C N G A R C C N G C N G C N G C N G C N C N G T N G C N G A R C C N G C N G C N 480
Qy 608 G A G C A C C T G T T G A G C T G A G C C T G T G C A G A G C A C C T T T G C A G C T G A G C C T G C T G C A 667
Db 481 G A R C N C N G T N G C N G A R C C N G C N G C N G A R C N C C N G T N G C N G A R C C N G C N G C N 540
Qy 668 G A G C A C C T G T G A G T G A G C C A G C T G C A G A G A C C T T C A C C A G C T G A G C C T G C T A C A 727
Db 541 G A R C N C N G T N G T N G A R C C N G C N G C N G A R C N S N C N G C N G A R C C N G C N A C N 600
Qy 728 G C C A A G C C T G T G C C C C A G A A C C T A C C C T T C T C C C T C T T G A A C A G G C A A A T C A G 784
Db 601 G C N A R C C N G C N G C N G A R C C N C A V C C N W S N C C N S N Y T N G A R C A G C N A A Y C A R 657

RESULT 7
US-09-962-832-225/c
; Sequence 225, Application US/09962832
; Patent No. US20020110821A1
; GENERAL INFORMATION:
; APPLICANT: Ebner, Reinhard
; TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Signatu
; FILE REFERENCE: 689290-74
; CURRENT APPLICATION NUMBER: US/09/962,832
; CURRENT FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US/60/235,077
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,280
; PRIOR FILING DATE: 2000-09-25
; NUMBER OF SEQ ID NOS: 259
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 225
; LENGTH: 2108
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-962-832-225

Query Match 9.7%; Score 98.2; DB 10; Length 2108;
Best Local Similarity 55.4%; Pred. No. 1.4e-12;
Matches 190; Conservative 0; Mismatches 153; Indels 0; Gaps 0;

Qy 428 G T T C C T C C T C C C T A G G G T T T C C C G T T T G C C C T C C T C A A G G T T T T T T C A G C A 487
Db 936 G C T G C T T C C T C C T G A T C A G G T G C T A C A G C T G C C C C T C C T G T T C C A G G T A C T T C A 877
Qy 488 G C T G C A C A C C C G T G C C C A C C A T T A T G A G C T G A G C C T G C T G C A G C T G C A C C T C T T A C A 547
Db 876 G C T G C C C C A T C T G C T C T G G A C C T C C A G C T G C C C C T C C T G G T G C T C C A G G T G C T T C A 817
Qy 548 G C C A C A C C T A G A G C T A G C C T G T G C A G G G C C C C T G T T G C A G C T A G C C T G C T G C A 607

Db 816 G C T G T C C C T C C T G C T C A G A G A C T C A G C T G C C C T C C T G C T C A G A G A C T C C A 757
Qy 608 G A G G C A C C T G T T G A G C T G A G C C T G C T G C A G A G A C C C T G T T G C A G C T G A G C C T G C T C A 667
Db 756 G C T G C C C C T C C T G C T G C T C A G A G A C T C A G C T G C C C C T C C T G C T G C T G G G A G C T C C A 697
Qy 668 G A G G C A C C T G T T G A G T G A G C C A G C T G C A G A G A A C C T T C A C A G C T G A G C C T G C T A C A 727
Db 696 G C T G C C C C T C C T G C T G C T G G A G C T C C A G C T G C C C C T C C T G C T G C T G G G A G C T C C A 637
Qy 728 G C C A A G C C T G C T G C C C C A G A A C C T C A C C C T T C C C T C C T C T T G 770
Db 636 G C T G C C C C T C C T G C T G C T G G A G C T C C A G C T G C C C C T C C T G 594

RESULT 8
US-09-874-062-2
; Sequence 2, Application US/09874062
; Patent No. US20020081607A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Four Disulfide Core Domain-Containing (FDCD) Polynucleotides,
; FILE REFERENCE: PT033P1
; CURRENT APPLICATION NUMBER: US/09/874,062
; CURRENT FILING DATE: 2001-06-06
; PRIOR APPLICATION NUMBER: PCT/US00/32462
; PRIOR FILING DATE: 2000-11-29
; PRIOR APPLICATION NUMBER: 60/168,229
; PRIOR FILING DATE: 1999-12-01
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 1101
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-874-062-2

Query Match 9.2%; Score 92.4; DB 10; Length 1101;
Best Local Similarity 57.7%; Pred. No. 2.1e-11;
Matches 165; Conservative 0; Mismatches 121; Indels 0; Gaps 0;

Qy 456 C G T T G T C C C T C C T C A A G G T T T T T C A G A G C T G C A G A C C C G C T G C C C C A C C T A T T G 515
Db 171 C A I T T C C A G T T C T G C A G C C T T C C T G C T A T C C A A C C C A G C T G C T A A A C C C A G C T G 230
Qy 516 C A G C T A G C C T G C T G A G C T G C A C C T T T A C A G C A C A C C C T G T A G C A G C T A G C C T G C T G 575
Db 231 C C T G A C C A C C T G C T G C A G A C C A C C T G C G C C C C A G C T G C A T T T C C A G T T G C T G 290
Qy 576 C A G G G C C C C T G T T G C A G C T A G C C T G C T G C A G A G C A C C T G T T G G A G C T A G C C T G C T G 635
Db 291 C A G G C C T T C C T G C T G T A T C T C C A G C T G C T G C A A C C C A G C T G C A G A C C A C C T G C T G 350
Qy 636 C A G A G G C A C C T G T T G C A G C T A G C C T G C T G C A G A G C A C C T G T T G G A G C T G G A G C A G C T G 695
Db 351 C G C C C C A G C T G C A T T T C C A G T T G C T G A G G C C T T C C T G C T A T C T C C A G C T G C T G 410
Qy 696 C A G A G A A C C T T C A C C A G C T A G C C T G C T A C A G C C A A C C C A G C C T G C T G C 741
Db 411 C A A A C C C A G C T G C T G C A G A C C A C C C T G C T G C C G C C C A G C T G C T G C 456

RESULT 9
US-09-874-062-2/c
; Sequence 2, Application US/09874062
; Patent No. US20020081607A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Four Disulfide Core Domain-Containing (FDCD) Polynucleotides,
; FILE REFERENCE: PT033P1
; CURRENT APPLICATION NUMBER: US/09/874,062

```
; CURRENT FILING DATE: 2001-06-06
; PRIOR APPLICATION NUMBER: PCT/US00/32462
; PRIOR FILING DATE: 2000-11-29
; PRIOR APPLICATION NUMBER: 60/168,229
; PRIOR FILING DATE: 1999-12-01
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 1101
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-874-062-2

Query Match      8.7%; Score 87.6; DB 10; Length 1101;
Best Local Similarity 57.2%; Pred. No. 2.4e-10;
Matches 159; Conservative 0; Mismatches 119; Indels 0; Gaps 0;

Qy 464 CCTCCTTCAAGTGTTCACGAGCTGCAGACGCCGCTGCCACCACTATTGACAGCTGAG 523
Db 384 CTGCAGCAACTGGAAATGACAGCAGCTGGGGCGGCAGCAGGTGGTCTGCAGCAGCTGGG 325

Qy 524 CTGCTGCAGCTGCACCTCTTACAGCCACACCTGTACAGCTGAGCCTGCTGCGAGGGCC 583
Db 324 TTTCAGCAGCTGGAGATACAGCAGGAAGCGCTGCAGCAACTGGAAATGCAGCAGCTGGG 265

Qy 584 CTTGTTGAGCTGAGCCTGTGTCAGAGGCGACCTGTTGGAGCTGAGCCTGCTGCGAGAGGCA 643
Db 264 CGCGCAGCAGTGGTCTGTCAGCAGCTGGTTCAGCAGCAGCTGGTTCGAGCAGCTGGA 205

Qy 644 CTTGTTGAGCTGAGCCTGTGTCAGAGGCGACCTGTTGGAGTGGAGCAGCTGCGAGAGAA 703
Db 204 GATACAGCAGCAAGCCCTGCAGCAACTGGAATGCAGCAGCTGGGGCGGCAGCAGGTGGT 145

Qy 704 CTTTACCAGCTGAGCCTGTGTCAGAGCAACGCTGCTGC 741
Db 144 CTGCAGCAGCTGCTGTGGCAGCAGCTGGGGCGGCAGC 107

RESULT 10
US-09-854-133-337/c
; Sequence 337, Application US/09854133
; Publication No. US20020183499A1
; GENERAL INFORMATION:
; APPLICANT: Lodes, Michael J.
; APPLICANT: Mohamath, Raodoh
; APPLICANT: Henderson, Robert A.
; APPLICANT: Benson, Darin R.
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; FILE REFERENCE: 210121.475C10
; CURRENT APPLICATION NUMBER: US/09/854,133
; CURRENT FILING DATE: 2001-05-11
; NUMBER OF SEQ ID NOS: 735
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 337
; LENGTH: 422
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(422)
; OTHER INFORMATION: n = A,T,C or G
US-09-854-133-337

Query Match      7.7%; Score 77.8; DB 9; Length 422;
Best Local Similarity 53.9%; Pred. No. 2.4e-08;
Matches 160; Conservative 0; Mismatches 137; Indels 0; Gaps 0;

Qy 481 TTCAGCAGCTGCAGACCCGCTGCCACCTATTGACAGCTGAGCCTGCTGCAGCTGCACC 540
Db 393 TGCAGCTCCGGCTGCAGGGGGTTCAGCTCCAGCTGCACCTCTCTGCTGGCGCTAGC 334

Qy 541 TCTTACAGCCACACCTGTAGCAGCTGAGCCTGCTGTCAGGGGGCCCTGTTGACAGCTGAGCC 600
Db 333 TCCACCGGGGTTCAGCTCCAGCTGCACCTCTCTGCTGCTCCAGCTGCTGCTGCTGCTCC 274

Qy 601 TGCTGCAGAGGCACTGTTGGAGCTGAGCCTGCTGCAGAGGCACTGTTGACAGCTGAGCC 660
Db 273 AGCTGTTGCTCCAGCTGCTGCAGCTCCTCTCGAGCTTTCAGCTGAGCTGCCGCTGCTCC 214

Qy 661 TGCTGCAGAGGCACTGTTGGAGTGGAGCGCAGCTGCAGAGGAACCTTTCACAGCTGAGCC 720
Db 213 TGCTGCGGTTCCAGCTCCTGCTGCCGCTCCAACTCCTGCTGCCGCTGCTGCCAGCTCCTGC 154

Qy 721 TGCTGCAGCAAGCCTGTGTCGCCCGCAGAACCTTCACCCCTTCTCCCTCTCTTGAACAGCC 777
Db 153 TCTGACCCCGAGGTCCACCGGCATGAGCTCCAGCTCCACCTCTTCTCCTCTCCGCGC 97
```

```
Qy 541 TCTTACAGCCACACCTGTAGCAGCTGAGCCTGCTGCAGGGGCCCTGTTGACAGCTGAGCC 600
Db 333 TCCACCGGGGTTCAGCTCCAGCTGCACCTCCTGCTCCAGCTGCTGCTGCTGCTGCTCC 274

Qy 601 TGCTGCAGAGGCACTGTTGGAGCTGAGCCTGCTGCAGAGGCACTGTTGACAGCTGAGCC 660
Db 273 AGCTGTTGCTCCAGCTGCTGCAGCTCCTCTGCTGCCGCTCCAACTCCTGCTGCCGCTGCTCC 214

Qy 661 TGCTGCAGCAAGCCTGTGTCGCCCGCAGAACCTTCACCCCTTCTCCCTCTCTTGAACAGCC 777
Db 153 TCTGACCCCGAGGTCCACCGGCATGAGCTCCAGCTCCACCTCTTCTCCTCTCCGCGC 97

RESULT 11
US-09-738-973-337/c
; Sequence 337, Application US/09738973
; Patent No. US20020110563A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Fling, Steven P.
; APPLICANT: Mohamath, Raodoh
; APPLICANT: Algate, Paul A.
; APPLICANT: Secrist, Heather
; APPLICANT: Indrias, Carol Yoseph
; APPLICANT: Benson, Darin R.
; APPLICANT: Elliot, Mark
; APPLICANT: Mannion, Jane
; APPLICANT: Kalos, Michael D.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; FILE REFERENCE: 210121.475C9
; CURRENT APPLICATION NUMBER: US/09/738,973
; CURRENT FILING DATE: 2000-12-14
; NUMBER OF SEQ ID NOS: 587
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 337
; LENGTH: 422
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(422)
; OTHER INFORMATION: n = A,T,C or G
US-09-738-973-337

Query Match      7.7%; Score 77.8; DB 10; Length 422;
Best Local Similarity 53.9%; Pred. No. 2.4e-08;
Matches 160; Conservative 0; Mismatches 137; Indels 0; Gaps 0;

Qy 481 TTCAGCAGCTGCAGCACCCTGCCACCTATTGACAGCTGAGCCTGCTGCAGCTGCACC 540
Db 393 TGCAGCTCCGGCTGCAGGGGGTTCAGCTCCAGCTGCACCTCTCTGCTGGCGCTAGC 334

Qy 541 TCTTACAGCCACACCTGTAGCAGCTGAGCCTGCTGTCAGGGGGCCCTGTTGACAGCTGAGCC 600
Db 333 TCCACCGGGGTTCAGCTCCAGCTGCACCTCTCTGCTGCTCCAGCTGCTGCTGCTGCTCC 274

Qy 601 TGCTGCAGAGGCACTGTTGGAGCTGAGCCTGCTGCAGAGGCACTGTTGACAGCTGAGCC 660
Db 273 AGCTGTTGCTCCAGCTGCTGCAGCTCCTCTCGAGCTTTCAGCTGAGCTGCCGCTGCTCC 214

Qy 661 TGCTGCAGAGGCACTGTTGGAGTGGAGCGCAGCTGCAGAGGAACCTTTCACAGCTGAGCC 720
Db 213 TGCTGCGGTTCCAGCTCCTGCTGCCGCTCCAACTCCTGCTGCCGCTGCTGCCAGCTCCTGC 154

Qy 721 TGCTGCAGCAAGCCTGTGTCGCCCGCAGAACCTTCACCCCTTCTCTCTCTTGAACAGCC 777
Db 153 TCTGACCCCGAGGTCCACCGGCATGAGCTCCAGCTCCACCTCTTCTCCTCTCCGCGC 97
```



```
Db 153 TCTGACCCAGGTCCACCGCATGAGCTCCAGCTCCACCTCTTCTCTCTCTCTGGGC 97
; Sequence 1, Application US/09922469
; Patent No. US2002017307A1
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; APPLICANT: Adler, David A.
; TITLE OF INVENTION: SECRETED SALIVARY ZSIC63 POLYPEPTIDE
; FILE REFERENCE: 97-71
; CURRENT APPLICATION NUMBER: US/09/922,469
; CURRENT FILING DATE: 2001-08-03
; PRIOR APPLICATION NUMBER: US 60/124,820
; PRIOR FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 1008
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (128)...(784)
US-09-922-469-1

Query Match 7.1%; Score 71.6; DB 9; Length 1008;
Best Local Similarity 56.3%; Pred. No. 8.4e-07;
Matches 134; Conservative 0; Mismatches 104; Indels 0; Gaps 0;

QY 463 CCTCTCTCAAGGTTTTTTTCAGCAGCTGCAGCACCCGCTGCCACCTATTGCGAGTGA 522
Db 700 CTCTGCAGCTGGCTCCACTCCACAGGTGCTCTGCAGCAGGCTCAGCTGCACAGGTGC 641
QY 523 GCCTGTGCGAGCTGACCTCTTTACAGCCACACCTGTGTAGCAGCTGAGCTGCTGCAGGGGC 582
Db 640 CTCTGCAGCAGGCTCAGCTCCACAGGTGCTCTGCAGCAGGCTCAGCTGCACAGGGGC 581
QY 583 CCTGTGTGCGAGCTGAGCTGCTGCAGAGGACCTGTTGGAGCTGAGCTGCTGCAGAGGC 642
Db 580 CCTGTGCGAGCTGAGCTGCTGCAGGTGGCTGTAGAGGTGCGAGCTGCAGAGGCTC 521
QY 643 ACCTGTGCGAGCTGAGCTGCTGCAGAGGACCTGTTGGAGTGGAGCGCAGCTGCAGAG 700
Db 520 AGCTGCAATAGTGGGCGAGCGGTGCTGCAGCTGCTGAAAAAACCTTGAAGGAGGG 463

RESULT 14
US-09-923-236-1/c
; Sequence 1, Application US/09923236
; Patent No. US20020090677A1
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; APPLICANT: Adler, David A.
; TITLE OF INVENTION: SECRETED SALIVARY ZSIC63 POLYPEPTIDE
; FILE REFERENCE: 97-71
; CURRENT APPLICATION NUMBER: US/09/923,236
; CURRENT FILING DATE: 2001-08-03
; PRIOR APPLICATION NUMBER: US 60/124,820
; PRIOR FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 1008
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (128)...(784)
US-09-923-236-1

Query Match 7.1%; Score 71.6; DB 10; Length 1008;
Best Local Similarity 56.3%; Pred. No. 8.4e-07;
Matches 134; Conservative 0; Mismatches 104; Indels 0; Gaps 0;

QY 463 CCTCTCTCAAGGTTTTTTTCAGCAGCTGCAGCACCCGCTGCCACCTATTGCGAGTGA 522
Db 700 CTCTGCAGCTGGCTCCACTCCACAGGTGCTCTGCAGCAGGCTCAGCTGCACAGGTGC 641
QY 523 GCCTGTGCGAGCTGCACCTCTTTACAGCCACACCTGTGTAGCAGCTGAGCTGCTGCAGGGGC 582
Db 640 CTCTGCAGCAGGCTCAGCTCCACAGGTGCTCTGCAGCAGGCTCAGCTGCACAGGGGC 581
QY 583 CCTGTGTGCGAGCTGAGCTGCTGCAGAGGACCTGTTGGAGCTGAGCTGCTGCAGAGGC 642
Db 580 CCTGTGCGAGCAGGCTCAGCTGCTGCAGGTGGCTGTAGAGGTGCGAGCTGCAGAGGCTC 521
QY 643 ACCTGTGCGAGCTGAGCTGCTGCAGAGGACCTGTTGGAGTGGAGCGCAGCTGCAGAG 700
Db 520 AGCTGCAATAGTGGGCGAGCGGTGCTGCAGCTGCTGAAAAAACCTTGAAGGAGGG 463

RESULT 15
US-09-822-846-491/c
; Sequence 491, Application US/09822846
; Publication No. US20030027139A1
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John M.
; APPLICANT: LaVallie, Edward R.
```

Wed Feb 12 16:04:17 2003

```
; APPLICANT: Collins-Racie, Lisa A.
; APPLICANT: Evans, Cheryl
; APPLICANT: Merberg, David
; APPLICANT: Treacy, Maurice
; APPLICANT: Agostino, Michael J.
; APPLICANT: Steininger II, Robert J.
; APPLICANT: Bowman, Michael R.
; APPLICANT: Spaulding, Vikki
; APPLICANT: Wong, Gordon G.
; APPLICANT: Clark, Hilary
; APPLICANT: Fechtel, Kim
; APPLICANT: Howes, Steven H.
; APPLICANT: Resnick, Richard J.
; APPLICANT: Gulukota, Kamalakara
; APPLICANT: Graham, James R.
; APPLICANT: Genetics Institute, Inc.
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS
; FILE REFERENCE: GIN 6400
; CURRENT APPLICATION NUMBER: US/09/822,846
; CURRENT FILING DATE: 2001-03-29
; PRIOR APPLICATION NUMBER: 60/195,605
; PRIOR FILING DATE: 2000-04-06
; NUMBER OF SEQ ID NOS: 629
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 491
; LENGTH: 2614
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-822-846-491

Query Match          7.1%; Score 71.6; DB 9; Length 2614;
Best Local Similarity 58.1%; Pred.No. 1.3e-06;
Matches 126; Conservative 0; Mismatches 91; Indels 0; Gaps 0;

Qy  466 TCCTTCAAGCTTTTTCAGCAGCTGCACGACCCCGCTGCCACCCTATTGCAGCTGAGCC 525
Db  2201 TGCTGTGGTGTGCTGTTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGC 2142

Qy  526 TGCTGCAGCTGCACCTTTACAGCCACCTGTAGCAGCTGAGCCCTGCAGGGGCCCC 585
Db  2141 TGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTG 2082

Qy  586 TGTTCAGCTGAGCCTGCTGCAGAGGACCCCTGTTGGAGCTGAGCCCTGCAGAGGACCC 645
Db  2081 ATGTGGTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGT 2022

Qy  646 TGTTCAGCTGAGCCTGCTGCAGAGGACCCCTGTTGGA 682
Db  2021 TGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1985
```

Search completed: February 12, 2003, 11:53:32
Job time : 62.3027 secs

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 12, 2003, 11:44:27 ; Search time 2461.58 Seconds
(without alignments)
10295.682 Million cell updates/sec

Title: US-09-923-236-1

Perfect score: 1008
Sequence: 1 agacagactaaagccat.....tagataatttagaccaatgg 1008

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 24791104 seqs, 12571243825 residues

Total number of hits satisfying chosen parameters: 49582208

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Pending_Patents_NA_Main.*

- 1: /cgn2_6/ptodata/1/pna/pctus_COMB.seq.*
- 2: /cgn2_6/ptodata/1/pna/US06_COMB.seq.*
- 3: /cgn2_6/ptodata/1/pna/US07_COMB.seq.*
- 4: /cgn2_6/ptodata/1/pna/US08_COMB.seq.*
- 5: /cgn2_6/ptodata/1/pna/US081_COMB.seq.*
- 6: /cgn2_6/ptodata/1/pna/US082_COMB.seq.*
- 7: /cgn2_6/ptodata/1/pna/US083_COMB.seq.*
- 8: /cgn2_6/ptodata/1/pna/US084_COMB.seq.*
- 9: /cgn2_6/ptodata/1/pna/US085_COMB.seq.*
- 10: /cgn2_6/ptodata/1/pna/US086_COMB.seq.*
- 11: /cgn2_6/ptodata/1/pna/US087_COMB.seq.*
- 12: /cgn2_6/ptodata/1/pna/US088_COMB.seq.*
- 13: /cgn2_6/ptodata/1/pna/US089_COMB.seq.*
- 14: /cgn2_6/ptodata/1/pna/US090_COMB.seq.*
- 15: /cgn2_6/ptodata/1/pna/US091_COMB.seq.*
- 16: /cgn2_6/ptodata/1/pna/US092_COMB.seq.*
- 17: /cgn2_6/ptodata/1/pna/US093_COMB.seq.*
- 18: /cgn2_6/ptodata/1/pna/US094_COMB.seq.*
- 19: /cgn2_6/ptodata/1/pna/US095A_COMB.seq.*
- 20: /cgn2_6/ptodata/1/pna/US095B_COMB.seq.*
- 21: /cgn2_6/ptodata/1/pna/US095C_COMB.seq.*
- 22: /cgn2_6/ptodata/1/pna/US095D_COMB.seq.*
- 23: /cgn2_6/ptodata/1/pna/US096A_COMB.seq.*
- 24: /cgn2_6/ptodata/1/pna/US096B_COMB.seq.*
- 25: /cgn2_6/ptodata/1/pna/US096C_COMB.seq.*
- 26: /cgn2_6/ptodata/1/pna/US096D_COMB.seq.*
- 27: /cgn2_6/ptodata/1/pna/US096E_COMB.seq.*
- 28: /cgn2_6/ptodata/1/pna/US097A_COMB.seq.*
- 29: /cgn2_6/ptodata/1/pna/US097B_COMB.seq.*
- 30: /cgn2_6/ptodata/1/pna/US097C_COMB.seq.*
- 31: /cgn2_6/ptodata/1/pna/US098A_COMB.seq.*
- 32: /cgn2_6/ptodata/1/pna/US098B_COMB.seq.*
- 33: /cgn2_6/ptodata/1/pna/US098C_COMB.seq.*
- 34: /cgn2_6/ptodata/1/pna/US099A_COMB.seq.*
- 35: /cgn2_6/ptodata/1/pna/US099B_COMB.seq.*
- 36: /cgn2_6/ptodata/1/pna/US099C_COMB.seq.*
- 37: /cgn2_6/ptodata/1/pna/US099D_COMB.seq.*
- 38: /cgn2_6/ptodata/1/pna/US100A_COMB.seq.*
- 39: /cgn2_6/ptodata/1/pna/US100B_COMB.seq.*
- 40: /cgn2_6/ptodata/1/pna/US101A_COMB.seq.*
- 41: /cgn2_6/ptodata/1/pna/US101B_COMB.seq.*
- 42: /cgn2_6/ptodata/1/pna/US102A_COMB.seq.*
- 43: /cgn2_6/ptodata/1/pna/US102B_COMB.seq.*

- 44: /cgn2_6/ptodata/1/pna/US6000_COMB.seq.*
- 45: /cgn2_6/ptodata/1/pna/US6001_COMB.seq.*
- 46: /cgn2_6/ptodata/1/pna/US6002_COMB.seq.*
- 47: /cgn2_6/ptodata/1/pna/US6003_COMB.seq.*
- 48: /cgn2_6/ptodata/1/pna/US6004_COMB.seq.*
- 49: /cgn2_6/ptodata/1/pna/US6005_COMB.seq.*
- 50: /cgn2_6/ptodata/1/pna/US6006_COMB.seq.*
- 51: /cgn2_6/ptodata/1/pna/US6007_COMB.seq.*
- 52: /cgn2_6/ptodata/1/pna/US6008_COMB.seq.*
- 53: /cgn2_6/ptodata/1/pna/US6009_COMB.seq.*
- 54: /cgn2_6/ptodata/1/pna/US6010_COMB.seq.*
- 55: /cgn2_6/ptodata/1/pna/US6011_COMB.seq.*
- 56: /cgn2_6/ptodata/1/pna/US6012_COMB.seq.*
- 57: /cgn2_6/ptodata/1/pna/US6013_COMB.seq.*
- 58: /cgn2_6/ptodata/1/pna/US6014_COMB.seq.*
- 59: /cgn2_6/ptodata/1/pna/US6015_COMB.seq.*
- 60: /cgn2_6/ptodata/1/pna/US6016_COMB.seq.*
- 61: /cgn2_6/ptodata/1/pna/US6017_COMB.seq.*
- 62: /cgn2_6/ptodata/1/pna/US6018_COMB.seq.*
- 63: /cgn2_6/ptodata/1/pna/US6019_COMB.seq.*
- 64: /cgn2_6/ptodata/1/pna/US6020_COMB.seq.*
- 65: /cgn2_6/ptodata/1/pna/US6021_COMB.seq.*
- 66: /cgn2_6/ptodata/1/pna/US6022_COMB.seq.*
- 67: /cgn2_6/ptodata/1/pna/US6023_COMB.seq.*
- 68: /cgn2_6/ptodata/1/pna/US6024_COMB.seq.*
- 69: /cgn2_6/ptodata/1/pna/US6025_COMB.seq.*
- 70: /cgn2_6/ptodata/1/pna/US6026_COMB.seq.*
- 71: /cgn2_6/ptodata/1/pna/US6027_COMB.seq.*
- 72: /cgn2_6/ptodata/1/pna/US6028_COMB.seq.*
- 73: /cgn2_6/ptodata/1/pna/US6029_COMB.seq.*
- 74: /cgn2_6/ptodata/1/pna/US6030_COMB.seq.*
- 75: /cgn2_6/ptodata/1/pna/US6031_COMB.seq.*
- 76: /cgn2_6/ptodata/1/pna/US6032_COMB.seq.*
- 77: /cgn2_6/ptodata/1/pna/US6033_COMB.seq.*
- 78: /cgn2_6/ptodata/1/pna/US6034_COMB.seq.*
- 79: /cgn2_6/ptodata/1/pna/US6035_COMB.seq.*
- 80: /cgn2_6/ptodata/1/pna/US6036_COMB.seq.*
- 81: /cgn2_6/ptodata/1/pna/US6037_COMB.seq.*
- 82: /cgn2_6/ptodata/1/pna/US6038_COMB.seq.*
- 83: /cgn2_6/ptodata/1/pna/US6039_COMB.seq.*
- 84: /cgn2_6/ptodata/1/pna/US6040_COMB.seq.*
- 85: /cgn2_6/ptodata/1/pna/US6041_COMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1008	100.0	1008	34	US-09-923-469-1
2	1008	100.0	1008	34	US-09-923-480-1
3	1008	100.0	1008	34	US-09-923-236-1
4	1001	99.3	1002	76	US-60-324-185-219
5	997	98.9	998	38	US-10-009-328-72
6	997	98.9	998	60	US-60-162-287-23
7	981.6	97.4	1325	1	PCT-US98-13813-9
8	981.6	97.4	1325	13	US-08-958-304-9
9	981.6	97.4	1325	22	US-09-746-783-105
10	967	95.9	979	22	US-09-585-715-149
11	967	95.9	979	60	US-60-167-410-37
12	944	93.7	956	22	US-09-585-715-10
13	944	93.7	956	57	US-60-137-261-9
14	931	92.4	961	1	PCT-US02-08123-144
15	931	92.4	961	1	PCT-US02-08277-108
16	931	92.4	961	1	PCT-US02-08278-143
17	896.4	88.9	959	1	PCT-US02-08123-695
18	896.4	88.9	959	1	PCT-US02-08277-471
19	896.4	88.9	959	1	PCT-US02-08278-666
20	896.4	88.9	959	1	PCT-US98-22376-41
21	896.4	88.9	959	16	US-09-296-622-41


```
Qy 486 CAGCTGCAGCACCGCTGCCACCTATTGACGTGAGCTGCTGCGAGCTGCACCTCTTAA 545
Db 481 CAGCTGCAGCACCGCTGCCACCTATTGACGTGAGCTGCTGCGAGCTGCACCTCTTAA 540
Qy 546 CAGCCACACCTGTAGCAGCTGAGCCTGCTGCAGGGGCCCTGTGTGACGTGAGCCTGCTG 605
Db 541 CAGCCACACCTGTAGCAGCTGAGCCTGCTGCAGGGGCCCTGTGTGACGTGAGCCTGCTG 600
Qy 606 CAGAGGCACCTGTTGGAGCTGAGCCTGCTGCAGAGGACACCTGTTTGCAGCTGAGCCTGCTG 665
Db 601 CAGAGGCACCTGTTGGAGCTGAGCCTGCTGCAGAGGACACCTGTTTGCAGCTGAGCCTGCTG 660
Qy 666 CAGAGGCACCTGTTGGAGCTGAGCAGCTGCAGAGGAACCTTCACAGCTGAGCCTGCTGA 725
Db 661 CAGAGGCACCTGTTGGAGCTGAGCAGCTGCAGAGGAACCTTCACAGCTGAGCCTGCTGA 720
Qy 726 CAGCCAAAGCCTGCTGCCCGAAGACCTCACCCCTTCTCTTGAACAGGCAAAATCAGT 785
Db 721 CAGCCAAAGCCTGCTGCCCGAAGACCTCACCCCTTCTCTTGAACAGGCAAAATCAGT 780
Qy 786 GAAATCTCTAGAGAGTACCATTGGGTTCAATTTCTATACTGATGCAGAAATAGTGAAT 845
Db 781 GAAATCTCTAGAGAGTACCATTGGGTTCAATTTCTATACTGATGCAGAAATAGTGAAT 840
Qy 846 CTACAAAAGTTTCTTCTTTTCCAAAGACTATTTTCATTCCTGTGTATTCAGAGTATTCA 905
Db 841 CTACAAAAGTTTCTTCTTTTCCAAAGACTATTTTCATTCCTGTGTATTCAGAGTATTCA 900
Qy 906 TCTCACTACATGATTTGTTGTGTGATTTTTCCTTGGACTTAATTTATATTTGAAAAA 965
Db 901 TCTCACTACATGATTTGTTGTGTGATTTTTCCTTGGACTTAATTTATATTTGAAAAA 960
Qy 966 CATTGATAATTAATAAATAAATAGATAATTAGAC 1002
Db 961 CATTGATAATTAATAAATAAATAGATAATTAGAC 997
```

RESULT 7

```
PCT-US98-13813-9
; Sequence 9, Application PC/TUS9813813
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John M.
; APPLICANT: LaVallie, Edward R.
; APPLICANT: Racie, Lisa A.
; APPLICANT: Treacy, Maurice
; APPLICANT: Spaulding, Vikki
; APPLICANT: Agostino, Michael J.
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
; TITLE OF INVENTION: ENCODING THEM
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: MA
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US98/13813
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Sprunger, Suzanne A.
; REGISTRATION NUMBER: 41,323
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 498-8284
; TELEFAX: (617) 876-5851
```

```
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1325 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
PCT-US98-13813-9

Query Match 97.4%; Score 981.6; DB 1; Length 1325;
Best Local Similarity 99.5%; Pred. No. 2.7e-185;
Matches 995; Conservative 0; Mismatches 4; Indels 1; Gaps 1;

Qy 10 AAAAAAGCCATGATTCTTTTCGTTCTCTCTAAAGAGAAAAATATATTTTAAAAATAC 69
Db 9 AAAAAAGCCATGATTCTTTTCGTTCTCTCTAAAGAGAAAAATATATTTTAAAAATAC 68
Qy 70 ATTGCGTATTTTCTAAACAATAAATTTATAGTGTATATTTATAGGTCAATCAAAAT 129
Db 69 ATTGCGTATTTTCTAAACAATAAATTTATAGTGTATATTTATAGGTCAATCAAAAT 128
Qy 130 GAAGCTTCTCCTTTGGGCTGCAATGATGGTCAACCACCTTCATCCATCTCTGAATATTC 189
Db 129 GAAGCTTCTCCTTTGGGCTGCAATGATGGTCAACCACCTTCATCCATCTCTGAATATTC 188
Qy 190 CTTCAATGGTGAGGATGACAATGACGATGGTCAACCACCTTCATCCATCTCTGAATATTC 249
Db 189 CTTCAATGGTGAGGATGACAATGACGATGGTCAACCACCTTCATCCATCTCTGAATATTC 248
Qy 250 TTATGCAATACGAAATTTACCACTCCTCTTTATATATGCCAGAGTAATACAGTCCCAAG 309
Db 249 TTATGCAATACGAAATTTACCACTCCTCTTTATATATGCCAGAGTAATACAGTCCCAAG 308
Qy 310 TTACCTGGGAATACTTACACTGACACAGGTTACCTTCGTATCCCTGGATTTCACTTC 369
Db 309 TTACCTGGGAATACTTACACTGACACAGGTTACCTTCGTATCCCTGGATTTCACTTC 368
Qy 370 TCCTGGATTCCTTATGTTATCACAATCGGTGTTTCCCTTAGCTACTCAGTTGAATGT 429
Db 369 TCCTGGATTCCTTATGTTATCACAATCGGTGTTTCCCTTAGCTACTCAGTTGAATGT 428
Qy 430 TCCTCCTCCTCCTCCTAGGGGTTTCCCGTTTGTCCCTCCTTCAAGGTTTTTTTCAGCAGC 489
Db 429 TCCTCCTCCTCCTCCTAGGGGTTTCCCGTTTGTCCCTCCTTCAAGGTTTTTTTCAGCAGC 488
Qy 490 TGCAGCACCGCTGCCCCACCTATTGACGTGAGCCTGCTGCAGCTGCACCTCTTACAGC 549
Db 489 TGCAGCACCGCTGCCCCACCTATTGACGTGAGCCTGCTGCAGCTGCACCTCTTACATC 548
Qy 550 CACACCTGTAGCAGCTGAGCCTGCTGCAGGGGCCCTGTTGACGTGAGCCTGCTGCAGA 609
Db 549 CACACCTGTAGCATCTGAGCCTGCTGCAGGGGCCCTGTTGACGTGAGCCTGCTGCAGA 608
Qy 610 GGCACCTGTTGGAGCTGAGCCTGCTGCAGAGCACCTGTTGACGTGAGCCTGCTGCAGA 669
Db 609 GGCACCTGTTGGAGCTGAGCCTGCTGCAGAGCACCTGTTGACGTGAGCCTGCTGCAGA 668
Qy 670 GGCACCTGTTGGAGCTGAGCCAGCTGCAGAGGAACCTTCACCAAGCTGAGCCTGCTACAGC 729
Db 669 GGCACCTGTTGGAGCTGAGCCAGCTGCAGAGGAACCTTCACCAAGCTGAGCCTGCTACAGC 728
Qy 730 CAGAGCTGCTGCCCGCAGAGCCTCACCTTCCTCCTCTCTTGAACAGGCAAAATCAGTGA 789
Db 729 CAAGAGCTGCTGCCCGCAGAGCCTCACCTTCCTCCTCTCTTGAACAGGCAAAATCAGTGA 788
Qy 790 TTCTCTAGAGAGTACCATTGGGTTCAATTTCTATACTGATGCAGAAATAGTGAATCTAC 849
Db 789 TTCTCTAGAGAGTACCATTGGGTTCAATTTCTATACTGATGCAGAAATAGTGAATCTAC 848
Qy 850 AAAAGTTTTTCTTTTCCAAAGACTATTTTCATTTCTGTTGATTCAGAGTATTCATCTC 909
Db 849 AAAAGTTTTTCTTTTCCAAAGACTATTTTCATTTCTGTTGATTCAGAGTATTCATCTC 908
```


Db 77 TTCTCCTTTGGCGCTGCATGTATGTATGTGCTTTTGAAGGAAGAGAGGGTTCCCTCTCA 136
QY 195 TTGGTCAGGATGACAAATGACAGTGGTCACCCACTTCATCCATCTCTGAATATTCCTTATG 254
Db 137 TTGGTCAGGATGACAAATGACAGTGGTCACCCACTTCATCCATCTCTGAATATTCCTTATG 196
QY 255 GCATACGGAATTTACACCTCCTCTTTATTTATTCGCCAGTGAATACAGTCCCGAGTTACC 314
Db 197 GCATACGGAATTTACACCTCCTCTTTATTTATTCGCCAGTGAATACAGTCCCGAGTTACC 256
QY 315 CTGGGAATTAATGACAGGTTACCTTCGTATCCCTCGATTCCTCGATTCCTCTG 374
Db 257 CTGGGAATTAATGACAGGTTACCTTCGTATCCCTCGATTCCTCTG 316
QY 375 GATTCCTCCTATGTCTATACATCCGTTTTCCTTACCTAGCTACTCAGTGAATGTCCTC 434
Db 317 GATTCCTCCTATGTCTATACATCCGTTTTCCTTACCTAGCTACTCAGTGAATGTCCTC 376
QY 435 CTCTCCTCCTAGGGGTTTCCCGTTTGTCCCTTCCCTTCAAGGTTTTCAGCAGCTGCAG 494
Db 377 CTCTCCTCCTAGGGGTTTCCCGTTTGTCCCTTCCCTTCAAGGTTTTCAGCAGCTGCAG 436
QY 495 CACCCGCTGCCCACTATGACAGGTTACCTTCGTATCCCTCGATTCCTCGATTCCTCTG 554
Db 437 CACCCGCTGCCCACTATGACAGGTTACCTTCGTATCCCTCGATTCCTCTG 496
QY 555 CTGTAGCAGCTGAGCCTGCTGCAGGGGCCCTGTTCGAGCTGAGCCTGCTGCAGAGGCAC 614
Db 497 CTGTAGCAGCTGAGCCTGCTGCAGGGGCCCTGTTCGAGCTGAGCCTGCTGCAGAGGCAC 556
QY 615 CTGTGGAGCTGAGCCTGCTGCAGAGGCACCTGTTCGAGCTGAGCCTGCTGCAGAGGCAC 674
Db 557 CTGTGGAGCTGAGCCTGCTGCAGAGGCACCTGTTCGAGCTGAGCCTGCTGCAGAGGCAC 616
QY 675 CTGTGGAGCTGAGCCTGCTGCAGAGGCACCTGTTCGAGCTGAGCCTGCTGCAGAGGCAC 734
Db 617 CTGTGGAGCTGAGCCTGCTGCAGAGGCACCTGTTCGAGCTGAGCCTGCTGCAGAGGCAC 676
QY 735 CTGTGGAGCTGAGCCTGCTGCAGAGGCACCTGTTCGAGCTGAGCCTGCTGCAGAGGCAC 794
Db 677 CTGTGGAGCTGAGCCTGCTGCAGAGGCACCTGTTCGAGCTGAGCCTGCTGCAGAGGCAC 736
QY 795 TAGAAGATACCATGGGTTCAATTTCTATCTACTGATGCAGAAATAAGTGAATTTCAAAAG 854
Db 737 TAGAAGATACCATGGGTTCAATTTCTATCTACTGATGCAGAAATAAGTGAATTTCAAAAG 796
QY 855 TTTTCTCTTTTCCAAAGACTATTTCAATTTCTGTTGATTCAGAGTATTCATCTCAGTAC 914
Db 797 TTTTCTCTTTTCCAAAGACTATTTCAATTTCTGTTGATTCAGAGTATTCATCTCAGTAC 856
QY 915 ATTGATTTGTTTGTGTTAGTTTTCCTTGGAGTTAAATTTATTTGAAAAAACATTGTATA 974
Db 857 ATTGATTTGTTTGTGTTAGTTTTCCTTGGAGTTAAATTTATTTGAAAAAACATTGTATA 916

RESULT 15
PCT-US02-08277-108
; Sequence 108, Application PC/TUS0208277
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc.
; TITLE OF INVENTION: Human Secreted Proteins
; FILE REFERENCE: PS907PCT
; CURRENT APPLICATION NUMBER: PCT/US02/08277
; PRIOR FILING DATE: 2002-03-19
; PRIOR APPLICATION NUMBER: US 60/331,287
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/306,171
; PRIOR FILING DATE: 2001-07-19
; PRIOR APPLICATION NUMBER: US 60/277,340

; PRIOR FILING DATE: 2001-03-21
; NUMBER OF SEQ ID NOS: 1357
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 108
; LENGTH: 961
; TYPE: DNA
; ORGANISM: Homo sapiens
PCT-US02-08277-108

Query Match 92.4%; Score 931; DB 1; Length 961;
Best Local Similarity 100.0%; Pred. No. 3e-175;
Matches 931; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 75 GTATTTCTTAAACAATAAATTTATAGTGTAAATATTCATAGGGTCAATCAAAATGAAGC 134
Db 17 GTATTTCTTAAACAATAAATTTATAGTGTAAATATTCATAGGGTCAATCAAAATGAAGC 76
QY 135 TTCTCCTTTGGGCTGCATTTATGTGTTGCTTTTGAAGGAAGAGACGGTTCCCTTTCA 194
Db 77 TTCTCCTTTGGGCTGCATTTATGTGTTGCTTTTGAAGGAAGAGACGGTTCCCTTTCA 136
QY 195 TTGTGAGGATGACAATGACAGTGGTCACCCACTTCATCCATCTCTGAATATTCCTTATG 254
Db 137 TTGTGAGGATGACAATGACAGTGGTCACCCACTTCATCCATCTCTGAATATTCCTTATG 196
QY 255 GCATACGGAATTTACCACTCCTCTTTATTTATTCGCCAGTGAATACAGTCCCGAGTTACC 314
Db 197 GCATACGGAATTTACCACTCCTCTTTATTTATTCGCCAGTGAATACAGTCCCGAGTTACC 256
QY 315 CTGGGAATTAATGACAGGTTACCTTCGTATCCCTCGATTCCTAACTTCCTCTG 374
Db 257 CTGGGAATTAATGACAGGTTACCTTCGTATCCCTCGATTCCTAACTTCCTCTG 316
QY 375 GATTCCTCCTATGTCTATACATCCGTTTTCCTTTAGCTACTCAGTGAATGTTCTCTC 434
Db 317 GATTCCTCCTATGTCTATACATCCGTTTTCCTTTAGCTACTCAGTGAATGTTCTCTC 376
QY 435 CTCTCCTCCTAGGGGTTTCCCGTTTGTCCCTCCTTCAAGGTTTTCAGCAGCTGCAG 494
Db 377 CTCTCCTCCTAGGGGTTTCCCGTTTGTCCCTCCTTCAAGGTTTTCAGCAGCTGCAG 436
QY 495 CACCCGCTGCCCACTATTCGAGCTGAGCCTGCTGCAGCTGCACCTCTTACAGCCACAC 554
Db 437 CACCCGCTGCCCACTATTCGAGCTGAGCCTGCTGCAGCTGCACCTCTTACAGCCACAC 496
QY 555 CTGTAGCAGCTGAGCCTGCTGCAGGGGCCCTGTTCGAGCTGAGCCTGCTGCAGAGGCAC 614
Db 497 CTGTAGCAGCTGAGCCTGCTGCAGGGGCCCTGTTCGAGCTGAGCCTGCTGCAGAGGCAC 556
QY 615 CTGTGGAGCTGAGCCTGCTGCAGAGGCACCTGTTCGAGCTGAGCCTGCTGCAGAGGCAC 674
Db 557 CTGTGGAGCTGAGCCTGCTGCAGAGGCACCTGTTCGAGCTGAGCCTGCTGCAGAGGCAC 616
QY 675 CTGTGGAGCTGAGCCTGCTGCAGAGGCACCTTTCAGCAGCTGAGCCTGCTGCAGCCAAAGC 734
Db 617 CTGTGGAGCTGAGCCTGCTGCAGAGGCACCTTTCAGCAGCTGAGCCTGCTGCAGCCAAAGC 676
QY 735 CTGCTGCCCCAGAACCTCAGCCCTTCTCCCTCTCTTGAACAGGCAATTCAGTGAATTTCTC 794
Db 677 CTGCTGCCCCAGAACCTCAGCCCTTCTCCCTCTCTTGAACAGGCAATTCAGTGAATTTCTC 736
QY 795 TAGAAGATACCATGGGTTCAATTTCTATCTACTGATGCAGAAATAAGTGAATTTCAAAAG 854
Db 737 TAGAAGATACCATGGGTTCAATTTCTATCTACTGATGCAGAAATAAGTGAATTTCAAAAG 796
QY 855 TTTTCTCTTTTCCAAAGACTATTTCAATTTCTGTTGATTCAGAGTATTCATCTCAGTAC 914
Db 797 TTTTCTCTTTTCCAAAGACTATTTCAATTTCTGTTGATTCAGAGTATTCATCTCAGTAC 856
QY 915 ATTGATTTGTTTGTGTTAGTTTTCCTTGGAGTTAAATTTATTTGAAAAAACATTGTATA 974
Db 857 ATTGATTTGTTTGTGTTAGTTTTCCTTGGAGTTAAATTTATTTGAAAAAACATTGTATA 916

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 12, 2003, 11:44:25 ; Search time 254.876 Seconds
(without alignments)

Title: US-09-923-236-1

Perfect score:

Sequence: 1 agacagactaaaaagccat.....tagataatttagaccaatgg 1008

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2768844 seqs, 741319522 residues

Total number of hits satisfying chosen parameters:	5537688
----------------------------------------------------	---------

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Pending_Patents_NA New: *

```

1: /cgn2_6/ptodata/1/pna/pct_NEW_OMB.seq.*
2: /cgn2_6/ptodata/1/pna/US07_NEW_OMB.seq.*
3: /cgn2_6/ptodata/1/pna/US06_NEW_OMB.seq.*
4: /cgn2_6/ptodata/1/pna/US08_NEW_OMB.seq.*
5: /cgn2_6/ptodata/1/pna/US09_NEW_OMB.seq.*
6: /cgn2_6/ptodata/1/pna/US10_NEW_OMB.seq.*
7: /cgn2_6/ptodata/1/pna/US10_NEW_OMB.seq.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	121.2	12.0	3489	6	US-10-194-046-1	Sequence 1, Appl
C 2	98.2	9.7	2108	6	US-10-240-435-1294	Sequence 1294, Ap
C 3	86.6	8.6	2250	6	US-10-144-771-24493	Sequence 24493, A
C 4	86	8.5	732	6	US-10-144-771-29644	Sequence 29644, A
C 5	85.2	8.5	1926	6	US-10-194-046-3	Sequence 3, Appl
C 6	85.2	8.5	8705	6	US-10-273-678-16	Sequence 16, Appl
C 7	85.2	8.5	8705	6	US-10-291-230-14	Sequence 14, Appl
C 8	85.2	8.5	8705	6	US-10-291-249-14	Sequence 14, Appl
C 9	85.2	8.5	10285	6	US-10-050-902-283	Sequence 283, App
C 10	85.2	8.5	16080	5	US-09-724-566A-48	Sequence 48, Appl
C 11	85.2	8.5	16080	5	US-09-471-6659A-48	Sequence 48, Appl
C 12	83.4	8.3	1044	6	US-10-144-771-31574	Sequence 31574, A
C 13	81.8	8.1	1118	6	US-10-144-771-33054	Sequence 33054, A
C 14	81.2	8.1	8043	1	PCT-US02-39133-40	Sequence 40, Appl
C 15	81	8.0	1954	6	US-10-144-771-4707	Sequence 4707, Ap
C 16	80.8	8.0	1954	6	US-10-144-771-4707	Sequence 4707, Ap
C 17	80	7.9	555	6	US-10-144-771-46530	Sequence 46530, A
C 18	80	7.9	1759	6	US-10-144-771-22760	Sequence 22760, A
C 19	79.8	7.9	729	6	US-10-144-771-31479	Sequence 31479, A
C 20	79.4	7.9	699	6	US-10-144-771-25548	Sequence 25548, A
C 21	77	7.6	961	6	US-10-144-771-31796	Sequence 31796, A
C 22	76.4	7.6	762	6	US-10-144-771-26386	Sequence 26386, A
C 23	76.2	7.6	1682	1	PCT-US02-18256-10	Sequence 10, Appl
C 24	75	7.4	819	6	US-10-144-771-21810	Sequence 21810, A
C 25	75	7.4	2282	1	PCT-US02-18256-7	Sequence 7, Appl
C 26	75	7.4	11507	5	US-09-724-676-12764	Sequence 12764, A

ALIGNMENTS

RESULT 1

US-10-194-046-1/c

; Sequence 1, Application US/10194046

GENERAL INFORMATION:

; APPLICANT: Robertson, Erle S.

; TITLE OF INVENTION: LANA Viral Protein Binding Sites

; FILE REFERENCE: UM-07142

; CURRENT APPLICATION NUMBER: US/10/194,046

; CURRENT FILING DATE: 2002-07-11

; PRIOR APPLICATION NUMBER: 09/410,399

; PRIOR FILING DATE: 1999-

; NUMBER OF SEQ ID NOS: 22

```

: SEC ID NO 1
: SOFTWARE: PatentIn version 3.1

```

```

; SEQ ID NO 1
: LENGTH: 3489

```

LENGTH: 3489
TYPE: DNA

TYPE: DNA
ORGANISM:

US-10-194-046-1
ORGANISM: Kaposi's sarcoma-associated herpesvirus

T
O
F
O
F
O
T
O
C
O

Query Match 12.0%; Score 121.2; DB 6; Length 3489;
Best Local Similarity 58.7%; Pred. No. 2.4e-15;
Matches 210; Conservative 0; Mismatches 148; Indels 0; Gaps 0;

QY	413	GCTACTCAGTTGAATTCCTTCCTCTCCCTCTCTAGGGTTTCCGTTGTGTCCTCTCTCA	472
Db	2188	GCTGCTCATCCTGCTGCTGCTCATCTGCTGCTGCTCATCTGCTGCTGCTGCTGCTCAT	2129
QY	473	AGTTTATTTTTCAGCAGCTGCAGACCGGTGCCACACTATTGAGCTGAGCCTGCTGCA	532
Db	2128	CCTGCTGCTCCTGCTCATCTGCTGCTGCTCATCTGCTGCTGCTGCTCATCTGCTGCT	2069
QY	533	GCTGACCTCTTTACAGCCACACTCTAGCAGTGAAGCTGCTGAGGGGCCCTGTTGCA	592
Db	2068	GCTCATCCTGCTGCTGCTCATCTGCTGCTCATCTGCTGCTCATCTGCTGCTGCTGCT	2009
QY	593	GCTGAGCTGTGTCAGAGGACCTGTTGGAGCTGAGCCTGCTGAGAGGACCTGTTGCA	652
Db	2008	GCTCATCTGCTGCTGCTCATCTGCTGCTCATCTGCTGCTGCTGCTGCTGCTGCTGCT	1949
QY	653	GCTGAGCTGCTGCAGAGGACCTGTTGGATGAGGACGCTGAGAGGAACCTTCACCA	712
Db	1948	GCTCATCTGCTGCTGCTCATCTGCTGCTGCTCATCTGCTGCTGCTCATCTGCTGCT	1889
QY	713	GCTGAGCTGTGTCAGAGGACGCTGCTGCCCGAGACCTCAACCTCTCCGCTCTCTTG	770
Db	1888	GCTCATCTGCTGCTGCTCATCTGCTGCTGCTCATCTGCTGCTGCTGCTGCTGCTGCTG	1831

RESULT 2

US-10-240-425-1294/C
; Sequence 1294, Application US/10240425
; GENERAL INFORMATION:
; APPLICANT: Williams, Amanda
; APPLICANT: Boland, Joseph F.
; APPLICANT: Lord, Reginald V.
; APPLICANT: Alvarez, Chris
; APPLICANT: Wetzel, Jon C.
; APPLICANT: Scherf, Uwe
; APPLICANT: Vockley, Joseph G.
; TITLE OF INVENTION: Gene Expression Profiles in Esophageal Tissue
; FILE REFERENCE: 44921-5026
; CURRENT APPLICATION NUMBER: US/10/240,425
; CURRENT FILING DATE: 2002-09-30
; PRIOR FILING DATE: 2001-03-28
; PRIOR APPLICATION NUMBER: US 60/193,446
; PRIOR FILING DATE: 2000-03-31
; NUMBER OF SEQ ID NOS: 1588
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1294
; LENGTH: 2108
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. M13903
US-10-240-425-1294

Query Match 9.7%; Score 98.2; DB 6; Length 2108;
Best Local Similarity 55.4%; Pred. No. 9.6e-11;
Matches 190; Conservative 0; Mismatches 153; Indels 0; Gaps 0;
QY 428 GTCTCTCTCTCCCTAGAGGGTTTCCCGTTGTCCTCTCTCAAGGTTTTTTCAGCA 487
DB 936 GCTGTCTCTCTGCTGATCAGGCTTCCAGCTGCCCTCTGCTGTTCCAGGTAATTCA 877
QY 488 GGTGACAGACCCCGCTGACCTATTGAGCTGAGCTGTGCTGACCTGCACCTCTTACA 547
DB 876 GCTGCCCATCTGCTCTCTGGGACCTCCAGCTGCCCTCTGCTGCTCCAGGTCCTCA 817
QY 548 GGCACACCTGTAGAGCTGAGCTGCTGAGGGGCCCTGTGCTGAGCTGAGCTGCTGCA 607
DB 816 GCTGTCTCTCTGCTGCTCAGAGAGCTCCAGCTGCCCTCTGCTGCTCAGAGAGCTCCA 757
QY 608 GAGGACCTGTTGGAGCTGAGCTGCTGCTGAGGACACCTGTTGAGCTGAGCTGCTGCA 667
DB 756 GCTGCCCTCTCTGCTGCTCAGAGAGCTCCAGCTGCCCTCTGCTGCTGCTGGAGCTCCA 697
QY 668 GAGGACCTGTTGGAGTGGAGCCAGCTGCTGAGAGAACCTTACACAGCTGAGCTGCTACA 727
DB 696 GCTGCCCTCTCTGCTGCTCTGGGAGCTCCAGCTGCCCTCTGCTGCTGCTGGAGCTCCA 637
QY 728 GCGAGCCTGCTGCTGCCAGCAACCTCACCTTCTCCCTCTCTTG 770
DB 636 GCTGCCCTCTCTGCTGCTGCTGGGAGCTCCAGCTGCCCTCTCTG 594

RESULT 3
US-10-144-771-24493
; Sequence 24493, Application US/10144771
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig
; TITLE OF INVENTION: HUMAN GENOME DISCOVERY SYSTEM AND USES THEREOF
; FILE REFERENCE: CL001321
; CURRENT APPLICATION NUMBER: US/10/144,771
; CURRENT FILING DATE: 2002-05-15
; NUMBER OF SEQ ID NOS: 47235
; SEQ ID NO 24493
; LENGTH: 2250
; TYPE: DNA
; ORGANISM: HUMAN
; FEATURE:
; NAME/KEY: misc_feature

; LOCATION: (1)...(2250)
; OTHER INFORMATION: n = A,T,C or G
US-10-144-771-24493
Query Match 8.6%; Score 86.6; DB 6; Length 2250;
Best Local Similarity 57.0%; Pred. No. 2.1e-08;
Matches 158; Conservative 0; Mismatches 119; Indels 0; Gaps 0;
QY 484 AGCAGCTGCAGCACCGCTGCCCCACCTATTGCTGAGCTGAGCTGCTGCTGAGCTGCTCT 543
DB 1332 AGCAGCTGCAGCCCAAGCAGCTGTGCCCAAGCAGCTGAGCCCAAGCAGCTGAGCCCAA 1391
QY 544 TACAGCCACACCTGTAGCAGCTGAGCTGCTGAGGGGCCCTGTTTGCAGCTGAGCTGCT 603
DB 1392 GGCAGCTGTGCCCCAAGCAGCTGAGCCCAAGCAGCTGTGCCCAAGCAGCTGAGCCCAA 1451
QY 604 TGCAGAGGCACCTGTTGGAGCTGAGCTGCTGAGAGGACCTGTTTGCAGCTGAGCTGCT 663
DB 1452 AGCAGCTGAGCCCGAGGAGCTGAGCCCAAGCAGCTGAGCCCAAGCAGCTGAGCCCGAG 1511
QY 664 TGCAGAGGCACCTGTTGGAGTGGAGCAGCTGCTGAGAGGAACCTTACACAGCTGAGCTGCT 723
DB 1512 GGCAGCTGAGCCCAAGCAGCTGAGCCCAAGCAGCTGTGCCCAAGCAGCTGAGCCCAA 1571
QY 724 TACAGCCCAAGCCTGCTGCCCCAGAACCTTCAACCTTCT 760
DB 1572 AGCAGCTGAGCCCAAGCAGCTGAGTCCCAAGGCGACT 1608

RESULT 4
US-10-144-771-29644/c
; Sequence 29644, Application US/10144771
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig
; TITLE OF INVENTION: HUMAN GENOME DISCOVERY SYSTEM AND USES THEREOF
; FILE REFERENCE: CL001321
; CURRENT APPLICATION NUMBER: US/10/144,771
; CURRENT FILING DATE: 2002-05-15
; NUMBER OF SEQ ID NOS: 47235
; SEQ ID NO 29644
; LENGTH: 732
; TYPE: DNA
; ORGANISM: HUMAN
US-10-144-771-29644

Query Match 8.5%; Score 86; DB 6; Length 732;
Best Local Similarity 54.8%; Pred. No. 2.3e-08;
Matches 170; Conservative 0; Mismatches 140; Indels 0; Gaps 0;
QY 451 TTTCCCTTTGCTCCCTCTTCAAGGTTTTTTTTCAGCAGCTGACAGCCCGCTGCCCAACC 510
DB 384 TCTCCCTCAAGCCCTGCAGTGGCTAGTCTTTCAGCAGCTAGTCTGTCAGCAGCTAGTCC 325
QY 511 TATTGACAGCTGAGCTGCTGCTGAGCTGACCTTTACAGCCACACCTGTAGCAGCTGAGCC 570
DB 324 TGCAGCGGCTAGTCTTTCAGCAGCTAGTCTGTCAGCGGCTAGTCTTTCAGCAGCTAGTCC 265
QY 571 TGCTGACAGGGGCCCTGTTGCTGAGCTGAGCTGCTGTCAGAGGACACCTGTTGAGCTGAGCC 630
DB 264 TGCAGCGGCTGCTCTGCTGAGCAGCTGCTCTTTCAGCAGCTAGTCTGTCAGCAGCTAGTCC 205
QY 631 TGCTGACAGGACACCTGTTGCTGAGCTGAGCTGCTGTCAGAGGACACCTGTTGAGTGGAGCC 690
DB 204 TTCAGCAGCTAGTCCCGCAGTGGCTAGTCTTTCAGCTAGTCTGCTGAGCGGCTAGTCC 145
QY 691 AGCTGACAGGAACCTTTCAGCAGCTGAGCTGCTGTCAGCCCAAGCCTGCTGCCCCAGAAC 750
DB 144 TGCAGCAGCTGCTCTTTCAGCAGCTAGTCCCGCAGCTAGTCTGCTGAGCGGCTAGTCC 85
QY 751 TCACCTCTCT 760
DB 84 TGAAGCGGCT 75

Qy	697	AGAGGAACCTTCACGAGTGTGATGCCTGTCTACAGCAAGCCTGTGCCCCAAGAACCCTCACCC	756
Db	10849	TCTGTCTGTGCCCTCCTGTCTGTGCCCTTCCTGTGCCCTCCTGTGCCCTCCTGTCTGTCTGC	10908
Qy	757	TTCTCCTCCTCT	766
Db	10909	CCCTCCTCCT	10918

```

RESULT 12
US-10-144-771-31574
; Sequence 31574, Application US/10144771
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig
; TITLE OF INVENTION: HUMAN GENOME DISCOVERY SYSTEM AND USES THEREOF
; FILE REFERENCE: CL001321
; CURRENT APPLICATION NUMBER: US/10/144, 771
; CURRENT FILING DATE: 2002-05-15
; NUMBER OF SEQ ID NOS: 47235
; SEQ ID NO 31574
; LENGTH: 1044
; TYPE: DNA
; ORGANISM: HUMAN
US-10-144-771-31574

```

```

RESULT 14
PCT-US02-39133-40
; Sequence 40, Application PC/TUS0239133
; GENERAL INFORMATION:
; APPLICANT: INCYTE GENOMICS, INC.
; APPLICANT: GRIFFIN Jennifer A.
; APPLICANT: RAMKUMAR, Jayalaxmi
; APPLICANT: EMERLING, Brooke M.
; APPLICANT: KABLE, Amy E.
; APPLICANT: ELLIOTT, Vicki S.
; APPLICANT: MARQUIS, Joseph P.
; APPLICANT: BAUGHN, Mariah R.
; APPLICANT: GORVAD, Ann E.
; APPLICANT: YUE, Henry
; APPLICANT: LEE, Ernestine A.
; APPLICANT: BECHA, Shanya D.
; APPLICANT: TANG, Y. Tom
; APPLICANT: TRAN, Uyen K.
; APPLICANT: SWARNAKAR, Anita
; APPLICANT: LEE, Sally
; APPLICANT: ISON, Craig H.
; APPLICANT: HAFALIA, April J.A.
; APPLICANT: TRAN, Bao
; APPLICANT: SPRAGUE, William W.
; APPLICANT: LEE, Soo Yeun
; APPLICANT: KHARE, Reena

```

[illegible]

```

RESULT 12
US-10-144-771-31574
; Sequence 31574, Application US/10144771
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig
; TITLE OF INVENTION: HUMAN GENOME DISCOVERY SYSTEM AND USES THEREOF
; FILE REFERENCE: CL001321
; CURRENT APPLICATION NUMBER: US/10/144,771
; CURRENT FILING DATE: 2002-05-15
; NUMBER OF SEQ ID NOS: 47235
; SEQ ID NO 31574
; LENGTH: 1044
; TYPE: DNA
; ORGANISM: HUMAN
; US-10-144-771-31574

```

RESULT 13
US-10-144-771-33054

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 12, 2003, 11:44:25 ; Search time 1475.98 Seconds
(without alignments)
11060.494 Million cell updates/sec

Title: US-09-923-236-1
Perfect score: 1008
Sequence: 1 agacagactaaaaagccat.....tagataatttagaccaatgg 1008

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues
Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*

- 1: em_estba:*
- 2: em_esthum:*
- 3: em_estin:*
- 4: em_estnu:*
- 5: em_estov:*
- 6: em_estpl:*
- 7: em_estro:*
- 8: em_hic:*
- 9: gb_est1:*
- 10: gb_est2:*
- 11: gb_hic:*
- 12: gb_est3:*
- 13: gb_est4:*
- 14: gb_est5:*
- 15: em_estfun:*
- 16: em_estom:*
- 17: gb_gss:*
- 18: em_gss_hum:*
- 19: em_gss_inv:*
- 20: em_gss_pln:*
- 21: em_gss_vrt:*
- 22: em_gss_fun:*
- 23: em_gss_mam:*
- 24: em_gss_mus:*
- 25: em_gss_other:*
- 26: em_gss_pro:*
- 27: em_gss_rtd:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	153	15.2	1302	11 AK015291	AK015291 Mus muscu
2	133	13.2	499	12 BF544627	BF544627 UI-R-B10-
c	3	131.2	13.0	492 17 AQ133079	AQ133079 HS_2188_A
	4	125.8	12.5	711 14 BM934267	BM934267 UI-M-CGOp
	5	108	10.7	605 17 AZ640388	AZ640388 1M0502E1
	6	108	10.7	718 17 AZ972907	AZ972907 2M0246M24

7	107.8	10.7	1070	17	CNS06PEX	AL409279 T3 end of
8	107.2	10.6	1024	17	CNS06NOT	AL407043 T3 end of
c	9	102.6	10.2	987	11 AK013560	AK013560 Mus muscu
	10	101.8	10.1	713	AZ346186	AZ346186 1M0081G06
	11	99.8	9.9	526	AZ312601	AZ312601 1M0028I04
	12	99.2	9.8	906	17 CNS02ILL	AL199074 Tetraodon
	13	98.8	9.8	943	17 CNS07BBW	AL437682 T7 end of
	14	98.2	9.7	600	12 BF006977	BF006977 1163138 A
	15	97.6	9.7	1003	17 CNS04OUZ	AL437682 T7 end of
	16	97	9.6	650	17 AZ337339	AL437682 T7 end of
c	17	96.2	9.5	725	14 EQ133889	EQ133889 UI-R-CNL-
	18	95.4	9.4	767	17 CNS01XFI	AL171639 Tetraodon
c	19	95	9.4	1007	13 BM563605	BM563605 AGENCOURT
	20	94.4	9.4	606	13 BJ314050	BJ314050 BJ314050
	21	92.6	9.2	368	17 BH244712	BH244712 ATZE14TF
c	22	91.4	9.1	551	9 A1676413	A1676413 etmEST016
	23	90.8	9.0	364	12 BG466926	BG466926 EESTed27
	24	90.4	9.0	443	17 FR0008252	292062 F.rubripes
	25	90	8.9	536	17 FR0036552	AL124061 Fugu rubr
c	26	89.6	8.9	599	13 BM601700	BM601700 170006870
c	27	89.2	8.8	2295	11 BC016304	BC016304 Homo sapi
c	28	89	8.8	599	17 BH244490	BH244490 ATZEC76TR
	29	89	8.8	920	11 AK004258	AK004258 Mus muscu
c	30	88.8	8.8	526	17 AZ312601	AZ312601 1M0028I04
	31	88.8	8.8	559	17 AZ661219	AZ661219 1M0539M10
	32	88.8	8.8	687	17 BH244710	BH244710 ATZE22TF
	33	88.8	8.8	801	17 BH314010	BH314010 CH230-86P
c	34	88.2	8.8	1114	14 BM811601	BM811601 AGENCOURT
c	35	87.4	8.7	920	11 AK004258	AK004258 Mus muscu
c	36	87.2	8.7	667	17 CNS03YKG	AL266285 Tetraodon
c	37	87.2	8.7	980	17 CNS01PHD	AL154817 Anopheles
	38	87	8.6	619	17 FR0013713	AL004959 F.rubripe
c	39	86.8	8.6	476	12 BF465646	BF465646 UI-M-CGOp
c	40	86.8	8.6	563	12 BE986054	BE986054 UI-M-CGOp
c	41	86.6	8.6	762	13 BJ313884	BJ313884 BJ313884
c	42	86	8.5	681	10 BB643958	BB643958 BB643958
c	43	85.6	8.5	644	12 BE995720	BE995720 UI-M-CGOp
	44	85.6	8.5	787	13 BJ282421	BJ282421 BJ282421
c	45	85.4	8.5	455	17 AQ928308	AQ928308 RPCI-23-2

ALIGNMENTS

RESULT 1	AK015291	AK015291	1302 bp	mRNA	linear	HTC 19-JAN-2002
LOCUS	Mus musculus adult male testis cDNA, RIKEN full-length enriched library, clone:4930432K09:hypothetical protein, full insert sequence.					
DEFINITION	AK015291					
ACCESSION	AK015291					
VERSION	AK015291.1	GI:12853570				
KEYWORDS	HTC; CAP trapper.					
SOURCE	Mus musculus (strain:C57BL/6J) adult male testis cDNA to mRNA, clone.lib:RIKEN full-length enriched mouse cDNA library					
ORGANISM	Mus musculus					
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.					
REFERENCE	1					
AUTHORS	Carninci, P., and Hayashizaki, Y.					
TITLE	High-efficiency full-length cDNA cloning					
JOURNAL	Meth. Enzymol. 303, 19-44 (1999)					
MEDLINE	99279253					
PUBMED	10349636					
REFERENCE	2					
AUTHORS	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.					
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes					
JOURNAL	Genome Res. 10 (10), 1617-1630 (2000)					
MEDLINE	20499374					
PUBMED	11042159					

Qy	577	AGGGCCCCCTGTTCCAGCTGAGCTGCTGCAGAGGCACCTGTTGGAGCTGAGCTGCTGC	636	
Db	661	AGGTGAGGGTCTGGTAGCTAGTTCCTGTAGACAAACCTATTTTAGGTCTGCCTCAGGC	720	
Qy	637	AGAGGCACCTGTTGCAGCTGAGCTGCTGCAGAGGC---	ACCTGTTGAGTGGAGCCAGC	693
Db	721	AGTCAAACTTGGACACCCGTGCCACCTCCGGACCCAAACCTCTTGACCCGAACTGC	780	
Qy	694	TGCAGAGAACTTTCACCACTGAGCTGCTGTACAGCCAAAGCTGCTGCCCCAGAACCTCA	753	
Db	781	TCTATCCCGCTTGGTGCACCTGAGCTGCCAGTCCAGTTCAGTTTGGTGCACCTGAGCCTGC	840	
Qy	754	CCC 756		
Db	841	CCC 843		
RESULT 2				
BF544627				
LOCUS	BF544627	499 bp	mRNA	
DEFINITION	UI-R-BT0-qi-c-05-0-UI.r1 UI-R-BT0 Rattus norvegicus cDNA clone			
ACCESSION	BF544627			
KEYWORDS	BF544627.1 GI:11635734			
SOURCE	EST.			
ORGANISM	Norway rat.			
	Rattus norvegicus			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
AUTHORS	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;			
TITLE	Rattus.			
	1 (bases 1 to 499)			
JOURNAL	Bonaldi,M.F., Lennon,G. and Soares,M.B.			
MEDLINE	Normalization and subtraction: two approaches to facilitate gene			
COMMENT	discovery			
	Genome Res. 6 (9), 791-806 (1996)			
	97044477			
	Contact: Soares, MB			
	Program for Rat Gene Discovery and Mapping			
	University of Iowa			
	451 Eckstein Medical Research Building Iowa City, IA 52242, USA			
	Tel: 319 335 8250			
	Fax: 319 335 9565			
	Email: msoares@blue.weeg.uiowa.edu			
	CDNA Library Preparation: M.B. Soares Lab Clone distribution:			
	clones will be available through Research Genetics (www.resgen.com)			
	This clone is also available through the I.M.A.G.E. Consortium at			
	LNL (info@image.lnl.gov). IMAGE ID= 1788839			
	Seq primer: M13 Forward.			
FEATURES	Location/Qualifiers			
source	1..499			
	/organism="Rattus norvegicus"			
	/strain="Sprague-Dawley"			
	/db_xref="taxon:10116"			
	/clone="UI-R-BT0-qi-c-05-0-UI"			
	/clone_lib="UI-R-BT0"			
	/dev_stage="adult"			
	/lab_host="DH10B (Life Technologies)"			
	/note="Vector: pT73D-Pac (Pharmacia) with a modified			
	polylinker; Site.1: Not I; Site.2: Eco RI; This library			
	(UI-R-BT0) consists of a mixture of individually tagged			
	normalized libraries constructed from rat hippocampus,			
	thalamus, mid-brain, medulla, corpus striatum, cerebral			
	cortex and testis. The tag used to identify the source			
	tissue is a string of 3-6 nucleotides present			
	Not I site and the oligo-dT track which allows			
	identification of the library of origin of a clone within			
	the mixture. This library was then subtracted using a			
	driver consisting of a mixture of all clones from UI-R-A0,			
	UI-R-A1, UI-R-E0, UI-R-E1, UI-R-C0, UI-R-C1, UI-R-C2 and			
	UI-R-C2p."			
BASE COUNT	123 a	148 c	80 g 148 t	
ORIGIN				

```

BASE COUNT      153 a      85 c      83 g      171 t
ORIGIN
    Query Match      13.0%; Score 131.2; DB 17; Length 492;
    Best Local Similarity 97.8%; Pred. No. 2.4e-15;
    Matches 133; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AGACAGCTAAAAAGCCATGTATCTTTCGTTTCTCTCTAAAGAAAGAAAAATATAATT 60
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 190 AGACAGCTAAATAAGCCATGTATCTTTCGTTTCTCTCTAAAGAAAGAAAAATATAATT 131
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 61 TAAAATACATTGCGTATTTCTAAACAATAAATTTATAGTGTAAATATTCATAGGCTC 120
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 130 TACAAATACATGCGTATTTCTAAACAATAAATTTATAGTGTAAATATTCATAGGCTC 71
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 121 AATCAAAATGAAGCTT 136
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 70 AATCAAAATGAAGCTT 55
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 4
BM934267
LOCUS
DEFINITION
    UI-M-CG0p-bib-g-02-0-UI.r1 NIH_BMAP_Ret4_S2 Mus musculus cDNA clone
    UI-M-CG0p-bib-g-02-0-UI 5', mRNA sequence.
ACCESSION
    BM934267
VERSION
    BM934267.1 GI:19393419
KEYWORDS
    EST.
SOURCE
    house mouse.
ORGANISM
    Mus musculus
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
    1 (bases 1 to 711)
AUTHORS
    Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE
    Normalization and subtraction: two approaches to facilitate gene
    discovery
JOURNAL
    Genome Res. 6 (9), 791-806 (1996)
MEDLINE
    9704477
COMMENT
    Contact: Chin, H
    National Institute of Mental Health
    6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
    20892-9643, USA
    Tel: 301 443 1706
    Fax: 301 443 9890
    Email: mEST@mail.nih.gov
    Tissue Procurement: Dr. Xin-Yuan Fu, Yale University School of
    Medicine
    cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
    cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
    DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
    Clone Distribution: Researchers may obtain clones from Research
    Genetics (www.resgen.com).
    Seq primer: M13 REVERSE.
    Location/Qualifiers
    1..711
    /organism="Mus musculus"
    /strain="C57BL/6J"
    /db_xref="taxon:10090"
    /clone="UI-M-CG0p-bib-g-02-0-UI"
    /clone_lib="NIH_BMAP_Ret4_S2"
    /lab_host="DH10B (Life Technologies)"
    /note="Vector: p773p-Pac (Pharmacia) with a modified
    polylinker. Site 1: Not I; Site 2: Eco RI; The
    NIH_BMAP_Ret4_S2 library is a subtracted library,
    ultimately derived from mouse retina tissue libraries at
    various stages of development. For a detailed description
    of the library from which this clone was derived, please
    visit our web site at brainest.eng.uiowa.edu. The tissue
    for this library was contributed by Dr. Xin-Yuan Fu, Yale
    University School of Medicine"
    173 a 215 c 116 g 204 t 3 others

BASE COUNT      173 a      215 c      116 g      204 t
ORIGIN
    Query Match      12.5%; Score 125.8; DB 14; Length 711;
    Best Local Similarity 53.8%; Pred. No. 2.5e-14;
    Matches 389; Conservative 0; Mismatches 274; Indels 60; Gaps 4;

QY 63 AAATACATTGCGTATTTCTAAACAATAAATTTATAGTGTAAATATTCATAGGCTCAA 122
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4 AAACACTCTGTGTATACATCCCTGAAATAAATCTACTATTTCATAGGCTCCA 63
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 123 TCRAAATGAAGCTTCCTCTTTGGCCTGCAATGTATGTGTGCTTTTGCAGGAAGAGAC 182
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 64 TCAAGATGAAGTTACTTCTTTGGCCTGCAATGTGTGTGCTTTTGCAGGAAGAGAGAC 123
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 183 GGTTCCTCCCTTCAT-----TGGTGAGGATGCAATATTCATTGGAATC 218
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 124 GTTATCCCTTCATTCACAAGAAATCGCCCTCATCCACTGAAGAGAGATTACTTTGGAATC 183
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 219 GTCACCCACTTCATCCATCTCTGAATATTCCTTATGGCATACGGAA----- 264
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 184 GCTACCCACTTAAACCATCTCTGAATATCCCTATGGCCTATGGAATGTAATTTGCCAC 243
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 265 ----TTTACCACCTCCCTCTTTATTATGCCCAGTGAATACAGTCCCAGTTACCTGGGA 320
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 244 CTTTGTGTTTGGCCTCCCTCTTAAACACTCAACAAGGAACGCCCATCACCAAATTCOCGGA 303
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 321 ATACTTACACTGACACAGGGTTTACCTTCGTATCCCTGGATTCTAAGTTCTCCTGGATTCC 380
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 304 ATTCTGAGCTTGAAGAGGGCTATCTCTATATCTTGGATTCCCACTTCTCCCAAAATTC 363
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 381 CCTATGCTATACATCGCGTGGTTTCCCTTAGCTACTCAAGTTGAATGTCTCCTCTCC 440
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 364 GCTATATATACCAAAACTATAAAGTACCCGCAGATGCGCCACTGAATGGTCTGCAATTG 423
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 441 CTCTAGGGGTTTCCCGTTTGTCCCTTCCTCAAGTTTTCAGCAGCTGCGAGCACCG 500
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 424 TTCC-----TCCTTCCTCCCTCCCGAGCCCTTTCCTCCAGGCCATATCCCT 468
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 501 CTGCCCCACCTATTGCGAGCTGAGCCTGCTGACGCTGCACCTCTTACAGGCACACCTGTAG 560
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 469 TTGTCATCCCTCCAAAGATTTCGTCTATCTCACCCTATGGACATGAGCCTGTAGCAGCTT 528
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 561 CAGCTGAGCCTGTGCGAGGGCCCTCTGTCAGCTGAGCCTGTGCGAGAGGACCTGTTG 620
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 529 CTGCTGTGCTGTGCTGAAGCGAGGCTTTGGTACCTGAGATCCCTGTAGATAAACAATATG 588
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 621 GAGCTGAGCCTGTGCGAGAGGCACCTGTTGACGCTGAGCCTGTGCGAGAGGC---ACCTG 677
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 589 CGGCTCTGCTCCGGTGATCAAACCTTGCAACCCCATGCCACCTCTGANACCANATCTC 648
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 678 TTGAGTGGAGCCAGCTGCGAGGAACCTTTCACACAGCTGAGCCTGTACAGCCCAAGCCTG 737
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 649 TTGCACCTCAACTGCTCCAGCCCACTTTTCTGTCACCCGAGCCTGTGCNAGTCCATTTTC 708
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 738 CTG 740
    ||
Db 709 TTG 711
    ||

RESULT 5
AZ640388
LOCUS
DEFINITION
    1M0502E21F Mouse 10kb plasmid UUGCIM library Mus musculus genomic
    clone UUGCIM0502E21 F, DNA sequence.
ACCESSION
    AZ640388
VERSION
    AZ640388.1 GI:11763788
KEYWORDS
    GSS.
SOURCE
    house mouse.
ORGANISM
    Mus musculus
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
    1 (bases 1 to 605)
AUTHORS
    Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamill,C.,
    Islam,H., Longacre,S., Mahmood,M., Meenen,E., Pedersen,T., Reilly

```

M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A. and Wright, D., Weiss, R.
Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
Unpublished (2000)
Contact: Robert B. Weiss
University of Utah Genome Center
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0502 row: E column: 21
Seq primer: CGTGTAAACGACGGCCAGT
Class: plasmid ends
High quality sequence stop: 605.
Location/Qualifiers
source
1. .605
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0502E21"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, Tl-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M. Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adapted DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi14732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."
BASE COUNT 13 a 253 c 136 g 203 t
ORIGIN
Query Match 10.7%; Score 108; DB 17; Length 605;
Best Local Similarity 53.9%; Pred. No. 6.8e-11;
Matches 222; Conservative 0; Mismatches 190; Indels 0; Gaps 0;
Qy 366 CTTCTCCGGAATCCCTATGCTATACATCCGCTGGTTTTCCCTAGCTACTACAGTTGA 425
Db 143 CTGCTCTCCCTCCCTAGGCGTTCCCGTTTCCCTCCCTCTAAGCTTTTTCAG 485
Qy 426 ATGTTCTCCCTCCCTAGGCGTTCCCGTTTCCCTCCCTCTAAGCTTTTTCAG 485
Db 203 CTACTGCTCTACTGCTCTACTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCC 262
Qy 486 CAGCTGACGACCCGCTGCCACCTATTGTCAGCTGAGCTGTCGTCGACCTCTTA 545
Db 263 CTGCTGCTTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGCC 322
Qy 546 CAGCCACACCTGTAGCAGCTGAGCTGTCGACGGGGCCCTGTTGTCAGCTGAGCCTGCTG 605
Db 323 CTGCTGCTCTCTGCTGCTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 382
Qy 606 CAGAGGCACCTGTTGGAGCTGAGCTGCTGTCGACGAGGACCTGTTGTCAGCTGAGCCTGCTG 665
Db 383 CTGCTGCTCTGTTGCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTC 442

Qy 666 CAGAGCACCTGTTGGAGTGAGCCAGCTGCAGAGGAACCTTCACAGCTGAGCCTGCTA 725
Db 443 CTGCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 502
Qy 726 CAGCCCAAGCCTGCTGCCCCAGAACCTTCACCCCTTCCTCCCTCTTGAACAGGC 777
Db 503 CTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGC 554
RESULT 6
AZ972907 718 bp DNA linear GSS 27-APR-2001
LOCUS 2M0246M24R Mouse 10kb plasmid UUGC2M library Mus musculus genomic
DEFINITION clone UUGC2M0246M24 R, DNA sequence.
ACCESSION AZ972907
VERSION AZ972907.1 GI:13844134
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 718)
Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C., Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A. and Wright, D., Weiss, R.
Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
Unpublished (2000)
Contact: Robert B. Weiss
University of Utah Genome Center
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0246 row: M column: 24
Seq primer: CACACGAGGAACAGCTATGACC
Class: plasmid ends
High quality sequence stop: 718.
Location/Qualifiers
Source
1. .718
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC2M0246M24"
/clone_lib="Mouse 10kb plasmid UUGC2M library"
/sex="Female"
/lab_host="E. coli strain XL10-Gold, Tl-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M. Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adapted DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi14732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."
BASE COUNT 43 a 291 c 144 g 240 t
ORIGIN

```
Query Match      10.7%; Score 108; DB 17; Length 718;
Best Local Similarity 55.3%; Pred. No. 6.7e-11;
Matches 210; Conservative 0; Mismatches 170; Indels 0; Gaps 0;

QY 398 CGTGGTTTCCCTAGCTACTCAGTTGATGTTCTCTCTCCCTCTAGGGGTTTCCCG 457
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 229 CCGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 288
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 458 TTTGTCCTCTCTCAAGGCTTTTTCAGCAGCTCAGCACCCGCTGCCACCTATTGCA 517
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 289 GCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 348
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 518 GCTGAGCCTGCTGAGCTGACCTCTTTACAGCCACACTCTTAGCAGCTGAGCCTGCGCA 577
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 349 GCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 408
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 578 GGGGCCCTCTTTGAGCTGAGCCTCTGCTGACAGGACACTGTTGAGCTGAGCCTGCGCA 637
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 409 GCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 468
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 638 GAGGCACCTGTTGAGCTGAGCCTGCTGCTGACAGGACACTGTTGAGTGAGCCAGCTGCA 697
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 469 GCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 528
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 698 GAGGAACCTTCACCACTGAGCCTGCTACAGCCAAAGCCTGCTGCCCAAGCCTCACCT 757
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 529 CCGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 588
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 758 TCTCCTCTCTTTGACAGGC 777
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 589 GCTCCTCTCTTGTCTCTGC 608
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 7
CNS06PEX
LOCUS      CNS06PEX      1070 bp      DNA      linear      GSS 06-JUL-2001
DEFINITION T3 end of clone AV00A014A04 of library AV00A from strain CBS 379 of
            Saccharomyces exiguus, genomic survey sequence.
ACCESSION  AL409279
VERSION     AL409279.1 GI:12176654
KEYWORDS   GSS.
SOURCE      Saccharomyces exiguus.
ORGANISM    Saccharomyces exiguus
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE   1 (bases 1 to 1070)
AUTHORS     Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
            Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
            de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
            Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
            Saurin,W., Tekala,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
            Wincker,P. and Weissenbach,J.
            Genomic exploration of the hemiascomycetous yeasts: 1. A set of
            yeast species for molecular evolution studies
            FEBS Lett. 487 (1), 3-12 (2000)
JOURNAL     20584711
MEDLINE     11152876
PUBMED      11152876
REFERENCE   2 (bases 1 to 1070)
AUTHORS     Bon,E., Neuveglise,C., Lepingle,A., Wincker,P., Artiguenave,F.,
            Galliardin,C. and Casaregola,S.
            Genomic exploration of the hemiascomycetous yeasts: 6.
            Saccharomyces exiguus
            FEBS Lett. 487 (1), 42-46 (2000)
JOURNAL     20584716
MEDLINE     11152881
PUBMED      11152881
REFERENCE   3 (bases 1 to 1070)
AUTHORS     Genoscope.
            Direct Submission
            Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage,
            2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
            seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
            This GSS is part of a random genomic sequencing program of thirteen
            yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
```

```
exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
5 kb were prepared and both extremities were sequenced. See
keywords for description of this sequence and for the sequence of
the other extremity of this insert.
FEATURES
    Location/Qualifiers
        1..1070
            /organism="Saccharomyces exiguus"
            /strain="CBS 379"
            /db_xref="taxon:34358"
            /clone="AV00A014A04"
            /clone_lib="AV00A"
            /note="end : T3"
        <11..>288
            /note="similar to Saccharomyces cerevisiae ORF YLR330w (
            CHS5 : chitin synthesis protein )
            1 putative frameshift(s)"
            /evidence=not_experimental
BASE COUNT  356 a  218 c  245 g  221 t      30 others
ORIGIN

Query Match      10.7%; Score 107.8; DB 17; Length 1070;
Best Local Similarity 61.3%; Pred. No. 7e-11;
Matches 160; Conservative 7; Mismatches 94; Indels 0; Gaps 0;

QY 491 GCAGCACCCTGCCCCACCTATTGCAGCTGAGCCCTGCTGCAGCTGCACCTCTTACAGCC 550
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 547 GAACCTGCTAMTGAGGAACCTGCTACTGAGGAACCTGCTACTGAGGAACCTGCTACTRAG 606
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 551 ACACCTGTAGCAGCTGAGCCTGCTGCAGGGGCCCTGTTGCAGCTGAGCCTGCTGCAGAG 610
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 607 GACMTGCTACTGAGGAACCTGCTACTGAGGAACCTGCTACTGAGGAACCTGCTACTGAG 666
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 611 GCACCTGTTGGAGCTGAGCCTGCTGCAGAGGACCTGTTGCAGCTGAGCCTGCTGCAGAG 670
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 667 GAACCTGMTACTGAGGAACCTGMTAMTGAGGAACCTGCTACTGAGGAACCTGCTGCCGAG 726
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 671 GCACCTGTTGGAGTGAGCAGCCTGCAGAGGACCTTCACCACTGAGCCTGCTGCAGCC 730
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 727 GAACCTGCTACTGAGGAACCTGMTACTGAGGAACCTGCTACTGAGGAACCTGCTACTGAC 786
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 731 AAGCCTGCTGCCCCAGAACT 751
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 787 GAACCACTGCCGAGGAACCT 807
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 8
CNS06NOT
LOCUS      CNS06NOT      1024 bp      DNA      linear      GSS 04-JUL-2001
DEFINITION T3 end of clone AU00AA015H08 of library AU00A from strain CBS 3082
            of Saccharomyces kluyveri, genomic survey sequence.
ACCESSION  AL407043
VERSION     AL407043.1 GI:12171906
KEYWORDS   GSS.
SOURCE      Saccharomyces kluyveri.
ORGANISM    Saccharomyces kluyveri.
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE   1 (bases 1 to 1024)
AUTHORS     Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
            Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
            de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
            Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
            Saurin,W., Tekala,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
            Wincker,P. and Weissenbach,J.
            Genomic exploration of the hemiascomycetous yeasts: 1. A set of
            yeast species for molecular evolution studies
            FEBS Lett. 487 (1), 3-12 (2000)
JOURNAL     20584711
MEDLINE     11152876
PUBMED      11152876
```


REFERENCE 2 (bases 1 to 906)
AUTHORS Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
Weissenbach,J.
TITLE Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 906)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (12-APR-2000)
COMMENT This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the Tetraodon nigroviridis
genome. For more information, please take a look at
http://www.genoscope.cns.fr/tetraodon.
FEATURES
source
1. .906
/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"
/clone="141F22"
/clone_lib="G"
/note="Genoscope sequence ID : COAG14DC11LPL1-end : T7"
BASE COUNT 155 a 271 c 259 g 216 t 5 others
ORIGIN
Query Match 9.8%; Score 99.2; DB 17; Length 906;
Best Local Similarity 58.1%; Pred. No. 3.2e-09;
Matches 173; Conservative 1; Mismatches 124; Indels 0; Gaps 0;
QY 444 CTAGGGGTTTCGCCGTTTGTCCCTCCTTCAAGGTTTTCACGAGCTGCAGCACCCGCTG 503
DB 444 GGGGGTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTC 503
QY 579 CCAGAGGCTCTTCGCCGCTGCACCAACAGTAGGAGCATGCTGCTGCTGCTGCTGCTG 638
DB 579 GGGGGTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTC 638
QY 504 CCCACCTATTGCAGCTGAGCCTGCTGCAGCTGCACCTTCCTACAGCCACACCTGTACAG 563
DB 504 TTTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCT 563
QY 639 CAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 698
DB 639 GGGGGTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTC 698
QY 564 CTGACGCTGCTGCAGGGGCCCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 623
DB 564 TTTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCT 623
QY 699 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 758
DB 699 TTTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCT 758
QY 624 CTGACGCTGCTGCAGAGCACCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 683
DB 624 TTTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCT 683
QY 759 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 818
DB 759 GGGGGTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTC 818
QY 684 TGGACGCTGCAGAGAACCTTTCACGCTGAGCCTGCTGCTGCTGCTGCTGCTGCTGCTG 741
DB 684 TTTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCT 741
QY 819 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 876
DB 819 TTTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCT 876
RESULT 13
CN507BBW 943 bp DNA linear GSS 08-JUL-2001
LOCUS T7 end of clone BC0AA010H05 of library BC0AA from strain CBS 767 of
DEFINITION Debaryomyces hansenii, genomic survey sequence.
ACCESSION AL437682
VERSION AL437682.1 GI:12221095
KEYWORDS GSS.
SOURCE Debaryomyces hansenii.
ORGANISM Debaryomyces hansenii.
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Debaryomyces.
REFERENCE 1 (bases 1 to 943)
AUTHORS Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
Saurin,W., Tekai,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
Wincker,P. and Weissenbach,J.
TITLE Genomic exploration of the hemiascomycetous yeasts: 1. A set of
yeast species for molecular evolution studies
JOURNAL FEBS Lett. 487 (1), 3-12 (2000)
MEDLINE 20584711

11152876
2 (bases 1 to 943)
REFERENCE Lepingle,A., Casaregola,S., Neuveglise,C., Bon,E., Nguyen,H.,
AUTHORS Artiguenave,F., Wincker,P. and Gaillardin,C.
TITLE Genomic exploration of the hemiascomycetous yeasts: 14.
JOURNAL Debaryomyces hansenii var. hansenii
MEDLINE FEBS Lett. 487 (1), 82-86 (2000)
PUBMED 20584724
11152889
3 (bases 1 to 943)
REFERENCE Genoscope.
AUTHORS Direct Submission
TITLE Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage,
JOURNAL 2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex - FRANCE (E-mail :
seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT This GSS is part of a random genomic sequencing program of thirteen
yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
5 kb were prepared and both extremities were sequenced. See
keywords for description of this sequence and for the sequence of
the other extremity of this insert.
FEATURES
Location/Qualifiers
1. .943
/organism="Debaryomyces hansenii"
/strain="CBS 767"
/variety="hansenii"
/db_xref="taxon:4959"
/clone="BC0AA010H05"
/clone_lib="BC0AA"
/note="end : T7"
BASE COUNT 232 a 204 c 283 g 218 t 6 others
ORIGIN
Query Match 9.8%; Score 98.8; DB 17; Length 943;
Best Local Similarity 60.4%; Pred. No. 3.8e-09;
Matches 163; Conservative 0; Mismatches 107; Indels 0; Gaps 0;
QY 481 TTGAGAGCTGCAGACACCGCTGCCCCACCTATTGAGCTGAGCTGCTGCTGCTGCTGCTG 540
DB 481 TTTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTC 540
QY 208 TTGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 267
DB 208 TTTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTC 267
QY 541 TCTTACAGCCACACCTGTAGCAGCTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 600
DB 541 TTTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTC 600
QY 268 GGTGGGGCTTTCACCTGCTGGAGCAGAACCGCTGGAGCTTACCTGCTGGAGTGTACC 327
DB 268 TTTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTC 327
QY 601 TGCTGAGAGGCACCTGTTGGAGCTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 660
DB 601 TTTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTC 660
QY 328 GGCTGGAGCTTTCACCTGCTGGAGCAGAACCGCTGGAGCTTACCTGCTGGAGTGTACC 387
DB 328 TTTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTC 387
QY 661 TGCTGAGAGGCACCTGTTGGAGCTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 720
DB 661 TTTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTC 720
QY 388 GGCTGGGGCTTTCACCTGCTGGAGTGTACCGCTGGGGCTTACCTGCTGGAGTGTACC 447
DB 388 TTTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTC 447
QY 721 TGCTACAGCCCAAGCTGCTGCCCCAGAAC 750
DB 721 TTTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTC 750
QY 448 AGCTGGGGCTTTCACCTGCTGGAGTGTACC 477
DB 448 TTTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTC 477
RESULT 14
BF006977 600 bp mRNA linear EST 06-OCT-2000
LOCUS 1163138 Amblyomma americanum adult Lambda 2ap Express Amblyomma
DEFINITION americanum cDNA, mRNA sequence.
ACCESSION BF006977
VERSION BF006977.1 GI:10707252
KEYWORDS EST.
SOURCE Amblyomma americanum.
ORGANISM Amblyomma americanum.
Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2003, 13:48:44 ; Search time 36 Seconds
(without alignments)
810.608 Million cell updates/sec

Title: US-09-923-236-2

Perfect score: 219

Sequence: 1 MKLLWACIVCAFAKRRF.....TAKPAPEPHSPSPSELANQ 219

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 908470 seqs, 133250620 residues

Word size : 0

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : A_Genesec_101002.*

- 1: /SID52/gcgdata/genesec/genesecp-emb1/AA1980.DAT.*
- 2: /SID52/gcgdata/genesec/genesecp-emb1/AA1981.DAT.*
- 3: /SID52/gcgdata/genesec/genesecp-emb1/AA1982.DAT.*
- 4: /SID52/gcgdata/genesec/genesecp-emb1/AA1983.DAT.*
- 5: /SID52/gcgdata/genesec/genesecp-emb1/AA1984.DAT.*
- 6: /SID52/gcgdata/genesec/genesecp-emb1/AA1985.DAT.*
- 7: /SID52/gcgdata/genesec/genesecp-emb1/AA1986.DAT.*
- 8: /SID52/gcgdata/genesec/genesecp-emb1/AA1987.DAT.*
- 9: /SID52/gcgdata/genesec/genesecp-emb1/AA1988.DAT.*
- 10: /SID52/gcgdata/genesec/genesecp-emb1/AA1989.DAT.*
- 11: /SID52/gcgdata/genesec/genesecp-emb1/AA1990.DAT.*
- 12: /SID52/gcgdata/genesec/genesecp-emb1/AA1991.DAT.*
- 13: /SID52/gcgdata/genesec/genesecp-emb1/AA1992.DAT.*
- 14: /SID52/gcgdata/genesec/genesecp-emb1/AA1993.DAT.*
- 15: /SID52/gcgdata/genesec/genesecp-emb1/AA1994.DAT.*
- 16: /SID52/gcgdata/genesec/genesecp-emb1/AA1995.DAT.*
- 17: /SID52/gcgdata/genesec/genesecp-emb1/AA1996.DAT.*
- 18: /SID52/gcgdata/genesec/genesecp-emb1/AA1997.DAT.*
- 19: /SID52/gcgdata/genesec/genesecp-emb1/AA1998.DAT.*
- 20: /SID52/gcgdata/genesec/genesecp-emb1/AA1999.DAT.*
- 21: /SID52/gcgdata/genesec/genesecp-emb1/AA2000.DAT.*
- 22: /SID52/gcgdata/genesec/genesecp-emb1/AA2001.DAT.*
- 23: /SID52/gcgdata/genesec/genesecp-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	219	100.0	219	22	AA860109 Human transport pr
2	219	100.0	219	23	AA074536 Human zsig63 polyp
3	150	68.5	221	20	AAV19472 Amino acid sequenc
4	140	63.9	219	20	AAW30653 Human secreted pro
5	9	4.1	75	22	AB864202 Drosophila melanog
6	9	4.1	322	17	AAW03688 Leishmania chagasi
7	9	4.1	728	22	AB865430 Drosophila melanog
8	9	4.1	1874	22	AA876532 Corynebacterium gl
9	9	4.1	2969	22	AAG90680 C glutamicum prote
10	8	3.7	65	21	AAG27598 Arabidopsis thalia

11	8	3.7	82	21	AA819327 Arabidopsis thalia
12	8	3.7	94	21	AA819326 Arabidopsis thalia
13	8	3.7	106	21	AA819325 Arabidopsis thalia
14	8	3.7	121	21	AA840893 Human ORFX ORF657
15	8	3.7	124	21	AA827837 Arabidopsis thalia
16	8	3.7	129	22	AA866380 Amino acid sequenc
17	8	3.7	132	20	AAV34678 Chlamydia pneumoni
18	8	3.7	139	21	AA854857 Arabidopsis thalia
19	8	3.7	145	21	AA832749 Eucalyptus grandis
20	8	3.7	145	21	AA807525 Arabidopsis thalia
21	8	3.7	145	21	AA845122 Arabidopsis thalia
22	8	3.7	145	21	AA845149 Arabidopsis thalia
23	8	3.7	181	21	AA806866 Arabidopsis thalia
24	8	3.7	181	21	AA810820 Arabidopsis thalia
25	8	3.7	185	21	AA854822 Arabidopsis thalia
26	8	3.7	193	21	AA806865 Arabidopsis thalia
27	8	3.7	193	21	AA810819 Arabidopsis thalia
28	8	3.7	205	21	AA806864 Arabidopsis thalia
29	8	3.7	205	21	AA810818 Arabidopsis thalia
30	8	3.7	328	21	AAV90267 Protein chimera Q
31	8	3.7	372	22	AA816669 Novel human diagno
32	8	3.7	412	21	AA802666 Protein chimera Q.
33	8	3.7	421	22	AB867110 Drosophila melanog
34	8	3.7	428	23	AA850963 Maize methyl Cpg b
35	8	3.7	493	22	AB870327 Drosophila melanog
36	8	3.7	511	22	AB828667 Peptide #1318 enco
37	8	3.7	511	22	AB833852 Peptide #1358 enco
38	8	3.7	511	22	AB819292 Protein #1291 enco
39	8	3.7	511	22	AA854816 Human brain expres
40	8	3.7	511	22	AA867022 Human bone marrow
41	8	3.7	511	22	AA814883 Peptide #1317 enco
42	8	3.7	511	22	AA827312 Peptide #1349 enco
43	8	3.7	511	22	AA802607 Peptide #1289 enco
44	8	3.7	511	23	AA836679 Human peptide enco
45	8	3.7	512	22	AB861369 Drosophila melanog

ALIGNMENTS

RESULT 1
AAB60109
ID AAB60109 standard; Protein; 219 AA.
XX
AC AAB60109;
XX
DF 28-MAR-2001 (first entry)
XX
DE Human transport protein TPPT-29.
XX
KW Human; transport protein; TPPT; transport disorder; metabolic disorder;
KW neurological disorder; cardiovascular disorder; reproductive disorder;
KW immune disorder; cancer.
XX
OS Homo sapiens.
XX
PN WO200078953-A2.
XX
PD 28-DEC-2000.
XX
PF 16-JUN-2000; 2000WO-US16668.
XX
PR 17-JUN-1999; 99US-0139923.
PR 10-AUG-1999; 99US-0148177.
PR 18-AUG-1999; 99US-0149357.
PR 28-OCT-1999; 99US-0162287.
PA (INCYTE) INCYTE GENOMICS INC.
XX
PI Lal P, Yang J, Yue H, Hillman JL, Tang YT, Bandman O, Burford N;
PI Baughn MR, Azimzai Y, Lu DAM, Au-young J, Patterson C;
XX
DW WPI; 2001-041424/05.

DR N-PSDB; AAF27729.

XX Isolated polypeptide with a human transport protein sequence is useful

PT for the diagnosis, prevention and treatment of disorders associated

PT with the immune, reproductive and cardiovascular systems -

XX

XX Claim 2; Page 130; 165pp; English.

XX The present invention provides the protein and coding sequences for 43

CC novel human transport proteins (designated TPTs). These can be used in

CC the diagnosis and treatment of transport, metabolic, neurological,

CC reproductive, cardiovascular and immune disorders, and cell proliferative

CC disorders such as cancer.

XX

SQ Sequence 219 AA;

Query Match 100.0%; Score 219; DB 22; Length 219;

Best Local Similarity 100.0%; Pred. No. 8.3e-201;

Matches 219; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLLWACIVCVAFARKRRPFIFIGEDDNDGHLHPSLNIPYGIKRLPPLPYRPNVTP 60

DB 1 MKLLWACIVCVAFARKRRPFIFIGEDDNDGHLHPSLNIPYGIKRLPPLPYRPNVTP 60

QY 61 SYPGNTYTDGLPSYFWILTSFGFYVYHIRGFPLATQLNVPLPRGPFVPPSRFFSA 120

DB 61 SYPGNTYTDGLPSYFWILTSFGFYVYHIRGFPLATQLNVPLPRGPFVPPSRFFSA 120

QY 121 AAAPAAPPTAAEPAAAPLTATPVAAEPAAAGAPVAAEPAAEPVGAEPAAEPVAAEPAA 180

DB 121 AAAPAAPPTAAEPAAAPLTATPVAAEPAAAGAPVAAEPAAEPVGAEPAAEPVAAEPAA 180

QY 181 EAPVGEPAAEPSPAEPATAKPAAPPEHPSPSLEQANQ 219

DB 181 EAPVGEPAAEPSPAEPATAKPAAPPEHPSPSLEQANQ 219

RESULT 2

AAU74536

ID AAU74536 standard; Protein; 219 AA.

XX

AC AAU74536;

XX

XX 23-APR-2002 (first entry)

DT

XX

XX Human zsig63 polypeptide.

DE

XX

XX Human; zsig63; chromosome 4q12-q13; salivary protein; antimicrobial;

KW microbial infection; tooth decay; periodontal disease; thrush; emphysema;

KW gastrointestinal disease; urinary tract infection; vaginal infection;

KW skin infection; epithelial wound; chronic tissue damage; cystic fibrosis;

KW acquired immunodeficiency syndrome; AIDS; lung infection; sarcoidosis;

KW chronic bronchitis; gene therapy; protein therapy.

XX

XX Homo sapiens.

OS

XX US6331413-B1.

PN

XX

XX 18-DEC-2001.

PD

XX

XX 17-MAR-2000; 2000US-0527345.

PF

XX

XX 17-MAR-1999; 99US-124820P.

XX

XX (ZYMO) ZYMOGENETICS INC.

PA

XX

XX Adler DA, Sheppard PO;

PI

XX

XX WPI; 2002-096707/13.

DR

XX

XX N-PSDB; AAS20591.

XX

XX Polynucleotides encoding salivary proteins useful as anti-microbial

PT agents -

PT

XX Claim 1; Column 49-52; 29pp; English.

PS

XX The invention relates to a polynucleotide derived from the 4q12-q13

CC region of human chromosome 4 and encoding a zsig63 polypeptide, a

CC secreted salivary protein with anti-microbial activity. Due to their

CC microbial activity, the sequences can be used in the study of microbial

CC infections, e.g. for recombinant production of anti-microbial proteins.

CC The sequences can be used in the treatment of tooth decay, periodontal

CC disease, thrush, gastrointestinal disease, urinary tract infections,

CC vaginal infections, skin infections, epithelial wounds, chronic tissue

CC damage, acquired immunodeficiency syndrome (AIDS), cystic fibrosis, lung

CC infections, sarcoidosis, emphysema and chronic bronchitis. This sequence

CC represents human zsig63.

XX

SQ Sequence 219 AA;

Query Match 100.0%; Score 219; DB 23; Length 219;

Best Local Similarity 100.0%; Pred. No. 8.3e-201;

Matches 219; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLLWACIVCVAFARKRRPFIFIGEDDNDGHLHPSLNIPYGIKRLPPLPYRPNVTP 60

DB 1 MKLLWACIVCVAFARKRRPFIFIGEDDNDGHLHPSLNIPYGIKRLPPLPYRPNVTP 60

QY 61 SYPGNTYTDGLPSYFWILTSFGFYVYHIRGFPLATQLNVPLPRGPFVPPSRFFSA 120

DB 61 SYPGNTYTDGLPSYFWILTSFGFYVYHIRGFPLATQLNVPLPRGPFVPPSRFFSA 120

QY 121 AAAPAAPPTAAEPAAAPLTATPVAAEPAAAGAPVAAEPAAEPVGAEPAAEPVAAEPAA 180

DB 121 AAAPAAPPTAAEPAAAPLTATPVAAEPAAAGAPVAAEPAAEPVGAEPAAEPVAAEPAA 180

QY 181 EAPVGEPAAEPSPAEPATAKPAAPPEHPSPSLEQANQ 219

DB 181 EAPVGEPAAEPSPAEPATAKPAAPPEHPSPSLEQANQ 219

RESULT 3

AAU19472

ID AAU19472 standard; Protein; 221 AA.

XX

AC AAU19472;

XX

XX 14-JUL-1999 (first entry)

DT

XX

XX Amino acid sequence of a human secreted protein.

DE

XX

XX Human secreted protein; cancer; tumour; neurodegenerative disorder;

KW developmental abnormality; fetal deficiency; blood disorder; leukemia;

KW immune system disease; autoimmune disease; hepatic disease;

KW renal disease; lymphoma; inflammation; allergy; ischemic shock;

KW Alzheimer's; cognitive disorder; schizophrenia; prostate disease;

KW obesity; osteoclast; osteoporosis; arthritis; malignancy; testes disease;

KW lung disease; thymus disease; digestive disorder; endocrine disorder;

KW infection; AIDS.

XX

XX Homo sapiens.

OS

XX

XX W09922243-A1.

PN

XX

XX 06-MAY-1999.

PD

XX

XX 23-OCT-1998; 98WO-US22376.

PF

XX

XX 24-OCT-1997; 97US-0063387.

XX

XX 24-OCT-1997; 97US-0062784.

PR

XX

XX 24-OCT-1997; 97US-0063088.

PR

XX

XX 24-OCT-1997; 97US-0063089.

PR

XX

XX 24-OCT-1997; 97US-0063090.

PR

XX

XX 24-OCT-1997; 97US-0063091.

PR

XX

XX 24-OCT-1997; 97US-0063092.

PR

XX

XX 24-OCT-1997; 97US-0063097.

PR 24-OCT-1997; 97US-0063098.
 PR 24-OCT-1997; 97US-0063099.
 PR 24-OCT-1997; 97US-0063100.
 PR 24-OCT-1997; 97US-0063101.
 PR 24-OCT-1997; 97US-0063109.
 PR 24-OCT-1997; 97US-0063110.
 PR 24-OCT-1997; 97US-0063111.
 PR 24-OCT-1997; 97US-0063148.
 PR 24-OCT-1997; 97US-0063386.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PA Brewer LA, Carter KC, Duan DR, Ebner R, Endress GA;
 PI Feng P, Florence C, Florence KA, Greene JM, Janat F;
 PI Kayw H, Lafleur DW, Moore PA, Ni J, Olsen HS, Rosen CA;
 PI Ruben SM, Shi Y, Soppet DR, Wei Y, Young P;
 XX
 DR WPI: 1999-303069/25.
 DR N-PSDB; AAX61352.
 XX
 PT New isolated human genes and the secreted polypeptides they encode
 XX
 PS Claim 11; Page 401-402; 546pp; English.
 XX
 CC The specification describes cDNA sequences (AAX61322-X61470) encoding
 CC human secreted proteins (AAY19442-Y19590). The polynucleotides and their
 CC corresponding secreted polypeptides are useful for preventing, treating
 CC or ameliorating medical conditions, e.g. by protein or gene therapy.
 CC Pathological conditions can also be diagnosed by determining the amount
 CC of the polypeptides in a sample or by determining the presence of
 CC mutations in the polynucleotides. Specific uses are described for each
 CC of the polynucleotides, based on which tissues they are most highly
 CC expressed in, and include developing products for the diagnosis or
 CC treatment of cancer, tumours, neurodegenerative disorders, developmental
 CC abnormalities and fetal deficiencies, blood disorders, leukemias,
 CC diseases of the immune system, autoimmune diseases, hepatic and renal
 CC diseases, lymphomas, inflammation, allergies, ischemic shock, Alzheimer's
 CC and cognitive disorders, schizophrenia, prostate diseases, obesity,
 CC disorders involving osteoclasts such as osteoporosis, arthritis or
 CC malignancies, diseases of testes, lung or thymus, digestive/endocrine
 CC disorders, infections and AIDS. The polypeptides are also useful for
 CC identifying their binding partners.
 XX
 SQ Sequence 221 AA;
 Query Match 68.5%; Score 150; DB 20; Length 221;
 Best Local Similarity 100.0%; Pred. No. 5.6e-135;
 Matches 150; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MKLLWACIVCVAFARRRPFPIGEDDDGHPHPSLNIPYGINLPPLPYRPVNTVP 60
 Db 1 MKLLWACIVCVAFARRRPFPIGEDDDGHPHPSLNIPYGINLPPLPYRPVNTVP 60
 QY 61 SYPGNTYTDGLPSYPIWLTSGFPVYHIRGFPLATQNLNPPPLPRGFPVPSRFFSA 120
 Db 61 SYPGNTYTDGLPSYPIWLTSGFPVYHIRGFPLATQNLNPPPLPRGFPVPSRFFSA 120
 QY 121 AAAPAAPPIAAEPAAAPLTAATPVAEPAA 150
 Db 121 AAAPAAPPIAAEPAAAPLTAATPVAEPAA 150
 RESULT 4
 AAW30653
 ID AAW30653 standard; Protein: 219 AA.
 XX
 AC AAW30653;
 XX
 DT 12-APR-1999 (first entry)
 XX
 DE Human secreted protein clone cpl16 1 protein.
 XX
 KW Human; secreted protein; nutritional activity; cytokine; vaccine;

KW cell proliferation; differentiation; immune stimulation; suppression;
 KW haematopoiesis regulation; tissue growth; activin; inhibin; chemotactic;
 KW chemokinetic; haemostatic; thrombolytic; anti-inflammatory; gene therapy;
 KW tumour invasion suppression; tumour inhibition.
 XX Homo sapiens.
 XX WO9501466-A1.
 PN 14-JAN-1999.
 PD
 XX
 XX 01-JUL-1998; 98WO-US13813.
 XX
 XX 27-OCT-1997; 97US-0958304.
 XX 02-JUL-1997; 97US-0887195.
 XX (GEMV) GENETICS INST INC.
 PA
 XX Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Racie LA;
 PI Spaulding V, Treacy M;
 XX
 DR WPI: 1999-105994/09.
 DR N-PSDB; AAV80740.
 XX
 PT New polynucleotides encoding secreted human proteins - are derived
 PT from human foetal brain, adult testes, adult brain, foetal kidney,
 PT adult salivary gland, or adult blood cDNA libraries, useful as, e.g.
 PT potential vaccines
 XX
 PS Claim 24; Page 71-72; 107pp; English.
 XX
 CC The present sequence represents a human secreted protein from clone
 CC cpl16 1, deposited as ATCC 98482. Human secreted protein clone
 CC polynucleotides and proteins are predicted to have biological
 CC activities which would make them suitable for treating, preventing or
 CC ameliorating medical conditions in humans and animals. Suggested
 CC activities include nutritional activity, cytokine and cell
 CC proliferation/differentiation activity, immune stimulating (e.g. as
 CC vaccines) or suppressing activity, haematopoiesis regulating activity,
 CC tissue growth activity, activin/inhibin activity, chemotactic/
 CC chemokinetic activity, haemostatic and thrombolytic activity, receptor/
 CC ligand activity, anti-inflammatory activity, cadherin/tumour invasion
 CC suppressor activity, and tumour inhibition activity. The polynucleotides
 CC are also stated to be useful for gene therapy.
 XX
 SQ Sequence 219 AA;
 Query Match 63.9%; Score 140; DB 20; Length 219;
 Best Local Similarity 100.0%; Pred. No. 1.9e-125;
 Matches 140; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MKLLWACIVCVAFARRRPFPIGEDDDGHPHPSLNIPYGINLPPLPYRPVNTVP 60
 Db 1 MKLLWACIVCVAFARRRPFPIGEDDDGHPHPSLNIPYGINLPPLPYRPVNTVP 60
 QY 61 SYPGNTYTDGLPSYPIWLTSGFPVYHIRGFPLATQNLNPPPLPRGFPVPSRFFSA 120
 Db 61 SYPGNTYTDGLPSYPIWLTSGFPVYHIRGFPLATQNLNPPPLPRGFPVPSRFFSA 120
 QY 121 AAAPAAPPIAAEPAAAPLTAATPVAEPAA 140
 Db 121 AAAPAAPPIAAEPAAAPLTAATPVAEPAA 140
 RESULT 5
 ABB64202
 ID ABB64202 standard; Protein: 75 AA.
 XX
 AC ABB64202;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Drosophila melanogaster polypeptide SEQ ID NO 19398.

XX Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical.
 XX
 OS Drosophila melanogaster.
 XX
 PN WO200171042-A2.
 XX
 PD 27-SEP-2001.
 XX
 PF 23-MAR-2001; 2001WO-US09231.
 XX
 PR 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 XX
 PA (PEKE) PE CORP NY.
 XX
 PI Venter JC, Adams M, Li PWD, Myers EW;
 XX WPI; 2001-656860/75.
 DR N-PSDB; ABL08305.
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 XX
 PS Disclosure; SEQ ID NO 19398; 21pp + Sequence Listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins
 CC (ABB57737-ABB72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 75 AA;
 Query Match 4.1%; Score 9; DB 22; Length 75;
 Best Local Similarity 100.0%; Pred. No. 0.76;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 120 AAAAPAAP 128
 Db 45 AAAAPAAP 53
 RESULT 6
 ID AAW03688 standard; Protein; 322 AA.
 AC AAW03688;
 XX
 DT 09-MAR-1997 (first entry)
 XX
 DE Leishmania chagasi acidic ribosomal antigen LcP0.
 XX
 KW Leishmania chagasi; acidic ribosomal antigen; LcP0;
 KW epitope; K39.
 XX
 OS Leishmania chagasi.
 XX
 PI Key Location/Qualifiers
 XX CDS 30..1202
 FT /*tag= a
 FT
 XX WO9633414-A2.
 PN
 XX 24-OCT-1996.
 PD

XX 19-APR-1996; 96WO-US05472.
 PF
 XX 21-APR-1995; 95US-0428414.
 PR
 XX (CORI-) CORIXA CORP.
 PA
 XX Reed SG;
 PI
 XX WPI; 1996-485884/48.
 DR N-PSDB; AAT42164.
 DR
 XX New Leishmania acidic ribosomal P-protein family polypeptide - used
 PT to develop prods. for diagnosis, detection and protection against
 PT Leishmania infections
 PT
 XX Disclosure; Page 29-32; 76pp; English.
 PS
 XX Compounds including polypeptides that contain at least an epitope of
 CC the L. chagasi acidic ribosomal antigen LcP0 are useful in a variety
 CC of immunoassays for detecting Leishmania infection. Portions of
 CC LcP0 (AAT42164) contg. at least the 17 C-terminal amino acids (AAT42165)
 CC have been found to generate a signal in an ELISA that is equivalent
 CC to that generated by the full length LcP0. A combination
 CC polypeptide may also be used, comprising an LcP0 epitope along with
 CC an epitope derived from the Leishmania K39 antigen (AAT42166), pref.
 CC the K39 repeat unit antigen having the sequence given in AAW03690.
 XX
 SQ Sequence 322 AA;
 Query Match 4.1%; Score 9; DB 17; Length 322;
 Best Local Similarity 100.0%; Pred. No. 2.6;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 130 AAEPAAAP 138
 Db 289 AAEPAAAP 297
 RESULT 7
 ID ABB65430 standard; Protein; 728 AA.
 XX
 AC ABB65430;
 XX
 DT 26-MAR-2002 (first entry)
 DT
 XX
 DE Drosophila melanogaster polypeptide SEQ ID NO 23082.
 XX
 KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical.
 XX
 OS Drosophila melanogaster.
 XX
 PN WO200171042-A2.
 PD
 PD 27-SEP-2001.
 XX
 PF 23-MAR-2001; 2001WO-US09231.
 XX
 PR 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 XX
 PA (PEKE) PE CORP NY.
 XX
 PI Venter JC, Adams M, Li PWD, Myers EW;
 XX WPI; 2001-656860/75.
 DR N-PSDB; ABL09533.
 DR
 XX New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 PT


```

XX Disclosure; SEQ ID NO 23082; 21pp + Sequence Listing; English.
PS
XX
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABU16176-ABU16051), expressed DNA
CC sequences (ABU16176-ABU16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX
SQ Sequence 728 AA;
  Query Match 4.1%; Score 9; DB 22; Length 728;
  Best Local Similarity 100.0%; Pred. No. 5.1;
  Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 119 SAAAPAAP 127
Db 418 SAAAPAAP 426
  |||||
  |||||

RESULT 8
AAB76532
ID AAB76532 standard; Protein; 1874 AA.
XX
AC AAB76532;
XX
DT 11-APR-2001 (first entry)
XX
DE Corynebacterium glutamicum MCT protein SEQ ID NO:46.
XX
KW Corynebacterium glutamicum; brevibacterium lactofermentum; MCT;
KW membrane construction and membrane transport protein; petroleum spill;
KW hydrocarbon degradation; gram positive aerobic bacterium; marker;
KW identification; microorganism; fine chemical production; transformation;
KW genome mapping; genetic engineering.
XX
XX
OS Corynebacterium glutamicum.
XX
XX
PN WO2000100805-A2.
XX
PD 04-JAN-2001.
XX
PF 23-JUN-2000; 2000WO-IB00926.
XX
XX
XX 25-JUN-1999; 99US-0141031.
PR 08-JUL-1999; 99DE-1031454.
PR 08-JUL-1999; 99DE-1031478.
PR 08-JUL-1999; 99DE-1031563.
PR 08-JUL-1999; 99DE-1032122.
PR 09-JUL-1999; 99DE-1032124.
PR 09-JUL-1999; 99DE-1032125.
PR 09-JUL-1999; 99DE-1032128.
PR 09-JUL-1999; 99DE-1032180.
PR 09-JUL-1999; 99DE-1032182.
PR 09-JUL-1999; 99DE-1032190.
PR 09-JUL-1999; 99DE-1032191.
PR 09-JUL-1999; 99DE-1032209.
PR 09-JUL-1999; 99DE-1032212.
PR 09-JUL-1999; 99DE-1032227.
PR 09-JUL-1999; 99DE-1032228.
PR 09-JUL-1999; 99DE-1032229.
PR 09-JUL-1999; 99DE-1032230.
PR 14-JUL-1999; 99DE-1032927.
PR 14-JUL-1999; 99DE-1033005.
PR 14-JUL-1999; 99DE-1033006.
PR 27-AUG-1999; 99DE-1040764.
PR 27-AUG-1999; 99DE-1040765.

```

```

PR 27-AUG-1999; 99DE-1040766.
PR 27-AUG-1999; 99DE-1040830.
PR 27-AUG-1999; 99DE-1040831.
PR 27-AUG-1999; 99DE-1040832.
PR 27-AUG-1999; 99DE-1040833.
PR 31-AUG-1999; 99DE-1041378.
PR 31-AUG-1999; 99DE-1041379.
PR 31-AUG-1999; 99DE-1041395.
PR 03-SEP-1999; 99DE-1042077.
PR 03-SEP-1999; 99DE-1042078.
PR 03-SEP-1999; 99DE-1042079.
XX 99DE-1042088.
XX
PA (BADI ) BASF AG.
XX
XX Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G;
XX
XX WPI; 2001-071486/08.
DR N-PSDB; AAF67765.
XX
XX Corynebacterium glutamicum nucleic acids encoding membrane construction
XX and membrane transport proteins or their portions, useful for typing or
XX identifying C. glutamicum or related bacteria, and as markers for
XX transformation.
XX
XX Claim 20; Page 202-207; 1119pp; English.
XX
XX AAF67743 to AAF68080 encode the Corynebacterium glutamicum membrane
XX construction and membrane transport (MCT) proteins given in AAB76510 to
XX AAB76847. The MCT nucleic acids and proteins are useful in the
XX identification of microorganisms which can be used to produce fine
XX chemicals, for modulating fine chemical production in C. glutamicum or
XX related bacteria (e.g. Brevibacterium lactofermentum), the typing or
XX identification of C. glutamicum or related bacteria, as reference points
XX for mapping C. glutamicum genome, and as markers for transformation.
XX AAF68082 and AAF68082 represent sequencing primers which are used in an
XX example from the present invention.
XX
XX Sequence 1874 AA;
  Query Match 4.1%; Score 9; DB 22; Length 1874;
  Best Local Similarity 100.0%; Pred. No. 11;
  Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 168 PAAEAPVAA 176
Db 568 PAAEAPVAA 576
  |||||
  |||||

RESULT 9
AAG90680
ID AAG90680 standard; Protein; 2969 AA.
XX
AC AAG90680;
XX
XX 26-SEP-2001 (first entry)
XX
XX C glutamicum protein fragment SEQ ID NO: 4434.
XX
XX Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
XX organic acid synthesis.
XX
XX Corynebacterium glutamicum.
XX
XX EP1108790-A2.
XX
XX 20-JUN-2001.
XX
XX 18-DEC-2000; 2000EP-0127688.
XX
XX 16-DEC-1999; 99JP-0377484.
XX 07-APR-2000; 2000JP-0159162.
XX 03-AUG-2000; 2000JP-0280988.

```

XX (KYOW) KYOWA HAKKO KOGYO KK.
PA Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX WPI: 2001-376931/40.
DR N-PSDB; AAH65899.
XX Novel polynucleotides derived from Coryneform bacteria, for identifying
PT mutation point of a gene, measuring expression of a gene, analysing
PT expression profile or pattern of a gene and identifying homologous gene
PT
XX
PS Claim 17; SEQ ID NO: 4434; 246pp + Sequence Listing; English.
XX
CC The present invention provides a number of nucleotide and protein
CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These
CC are useful for identifying the mutation point of a gene derived from a
CC mutant of coryneform bacterium, measuring expression amount and
CC analysing the expression profile or expression pattern of a gene derived
CC from Coryneform bacterium, and identifying a homologue of a gene derived
CC from coryneform bacterium. Coryneform bacteria are useful for producing
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC particularly L-lysine. The present sequence is a protein described
CC in the exemplification of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC European Patent Office.
XX
SQ Sequence 2969 AA;

Query Match 4.1%; Score 9; DB 22; Length 2969;
Best Local Smilarity 100.0%; Pred. No.16;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 168 PAAEAPVAA 176
Db 1663 PAAEAPVAA 1671
|||||||

RESULT 10
AAG27598
ID AAG27598 standard; Protein; 65 AA.
XX
AC AAG27598;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 32499.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-JUL-1999; 99US-0143334.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144325.
PR 16-JUL-1999; 99US-0144086.
PR 16-JUL-1999; 99US-0144085.
PR 15-JUL-1999; 99US-0143624.
PR 14-JUL-1999; 99US-0143542.
PR 13-JUL-1999; 99US-0142977.
PR 12-JUL-1999; 99US-0142920.
PR 09-JUL-1999; 99US-0142803.
PR 08-JUL-1999; 99US-0142390.
PR 06-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 01-JUL-1999; 99US-0141842.
PR 30-JUN-1999; 99US-0141287.
PR 29-JUN-1999; 99US-0140991.
PR 28-JUN-1999; 99US-0140895.
PR 24-JUN-1999; 99US-0140354.
PR 23-JUN-1999; 99US-0140353.
PR 22-JUN-1999; 99US-0139899.
PR 21-JUN-1999; 99US-0139817.
PR 18-JUN-1999; 99US-0139763.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139454.
PR 17-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 16-JUN-1999; 99US-0139452.
PR 14-JUN-1999; 99US-0139119.
PR 10-JUN-1999; 99US-0138847.
PR 10-JUN-1999; 99US-0138540.
PR 08-JUN-1999; 99US-0138094.
PR 07-JUN-1999; 99US-0137724.
PR 04-JUN-1999; 99US-0137502.
PR 03-JUN-1999; 99US-0137528.
PR 01-JUN-1999; 99US-0137222.
PR 27-MAY-1999; 99US-0136392.
PR 25-MAY-1999; 99US-0136021.
PR 24-MAY-1999; 99US-0135629.
PR 21-MAY-1999; 99US-0135353.
PR 20-MAY-1999; 99US-0135124.
PR 19-MAY-1999; 99US-0134941.
PR 18-MAY-1999; 99US-0134768.
PR 14-MAY-1999; 99US-0134570.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134218.
PR 11-MAY-1999; 99US-0134256.
PR 11-MAY-1999; 99US-0134263.
PR 08-MAY-1999; 99US-0134287.
PR 06-MAY-1999; 99US-0134286.
PR 05-MAY-1999; 99US-0134285.
PR 04-MAY-1999; 99US-0134284.
PR 30-APR-1999; 99US-0132407.
PR 30-APR-1999; 99US-0132048.
PR 28-APR-1999; 99US-0131449.
PR 23-APR-1999; 99US-0130891.
PR 23-APR-1999; 99US-0130510.
PR 21-APR-1999; 99US-0130077.
PR 19-APR-1999; 99US-0130077.

PR 19-JUL-1999;	99US-0144335.
PR 20-JUL-1999;	99US-0144352.
PR 20-JUL-1999;	99US-0144632.
PR 20-JUL-1999;	99US-0144884.
PR 21-JUL-1999;	99US-0144814.
PR 21-JUL-1999;	99US-0145086.
PR 21-JUL-1999;	99US-0145088.
PR 22-JUL-1999;	99US-0145087.
PR 22-JUL-1999;	99US-0145085.
PR 22-JUL-1999;	99US-0145089.
PR 22-JUL-1999;	99US-0145192.
PR 23-JUL-1999;	99US-0145145.
PR 23-JUL-1999;	99US-0145218.
PR 23-JUL-1999;	99US-0145224.
PR 26-JUL-1999;	99US-0145276.
PR 27-JUL-1999;	99US-0145913.
PR 27-JUL-1999;	99US-0145918.
PR 27-JUL-1999;	99US-0145919.
PR 28-JUL-1999;	99US-0145951.
PR 02-AUG-1999;	99US-0146386.
PR 02-AUG-1999;	99US-0146388.
PR 02-AUG-1999;	99US-0146389.
PR 03-AUG-1999;	99US-0147038.
PR 04-AUG-1999;	99US-0147204.
PR 04-AUG-1999;	99US-0147302.
PR 05-AUG-1999;	99US-0147192.
PR 05-AUG-1999;	99US-0147260.
PR 06-AUG-1999;	99US-0147303.
PR 06-AUG-1999;	99US-0147416.
PR 09-AUG-1999;	99US-0147493.
PR 09-AUG-1999;	99US-0147935.
PR 10-AUG-1999;	99US-0148171.
PR 11-AUG-1999;	99US-0148319.
PR 12-AUG-1999;	99US-0148341.
PR 13-AUG-1999;	99US-0148565.
PR 13-AUG-1999;	99US-0148684.
PR 16-AUG-1999;	99US-0149368.
PR 17-AUG-1999;	99US-0149175.
PR 18-AUG-1999;	99US-0149426.
PR 20-AUG-1999;	99US-0149722.
PR 20-AUG-1999;	99US-0149723.
PR 20-AUG-1999;	99US-0149929.
PR 23-AUG-1999;	99US-0149902.
PR 23-AUG-1999;	99US-0149930.
PR 25-AUG-1999;	99US-0150566.
PR 26-AUG-1999;	99US-0150894.
PR 27-AUG-1999;	99US-0151065.
PR 27-AUG-1999;	99US-0151066.
PR 27-AUG-1999;	99US-0151080.
PR 30-AUG-1999;	99US-0151303.
PR 31-AUG-1999;	99US-0151438.
PR 01-SEP-1999;	99US-0151930.
PR 07-SEP-1999;	99US-0152363.
PR 10-SEP-1999;	99US-0153070.
PR 13-SEP-1999;	99US-0153758.
PR 15-SEP-1999;	99US-0154018.
PR 16-SEP-1999;	99US-0154039.
PR 20-SEP-1999;	99US-0154779.
PR 22-SEP-1999;	99US-0155139.
PR 23-SEP-1999;	99US-0155486.
PR 24-SEP-1999;	99US-0155659.
PR 28-SEP-1999;	99US-0156458.
PR 29-SEP-1999;	99US-0156596.
PR 04-OCT-1999;	99US-0157117.
PR 05-OCT-1999;	99US-0157753.
PR 06-OCT-1999;	99US-0157865.
PR 07-OCT-1999;	99US-0158029.
PR 08-OCT-1999;	99US-0158232.
PR 12-OCT-1999;	99US-0158369.
PR 13-OCT-1999;	99US-0159293.
PR 13-OCT-1999;	99US-0159294.
PR 13-OCT-1999;	99US-0159295.
PR 14-OCT-1999;	99US-0159329.
PR 14-OCT-1999;	99US-0159330.
PR 14-OCT-1999;	99US-0159331.
PR 14-OCT-1999;	99US-0159637.
PR 18-OCT-1999;	99US-0159584.
PR 21-OCT-1999;	99US-0160741.
PR 21-OCT-1999;	99US-0160767.
PR 21-OCT-1999;	99US-0160768.
PR 21-OCT-1999;	99US-0160770.
PR 21-OCT-1999;	99US-0160814.
PR 22-OCT-1999;	99US-0160815.
PR 22-OCT-1999;	99US-0160980.
PR 22-OCT-1999;	99US-0160981.
PR 22-OCT-1999;	99US-0160989.
PR 25-OCT-1999;	99US-0161404.
PR 25-OCT-1999;	99US-0161405.
PR 25-OCT-1999;	99US-0161406.
PR 26-OCT-1999;	99US-0161359.
PR 26-OCT-1999;	99US-0161360.
PR 26-OCT-1999;	99US-0161361.
PR 28-OCT-1999;	99US-0161920.
PR 28-OCT-1999;	99US-0161992.
PR 28-OCT-1999;	99US-0161993.
PR 29-OCT-1999;	99US-0162142.
Query Match 3.7%; Score 8; DB 21; Length 65;	
Best Local Similarity 100.0%; Pred. No. 6.1;	
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY 130 AAEPAAAA 137	
Db 19 AAEPAANA 26	
RESULT 11	
AAG19327	
ID AAG19327 standard; Protein; 82 AA.	
XX	
AC AAG19327;	
XX	
DT 17-OCT-2000 (first entry)	
XX	
DE Arabidopsis thaliana protein fragment SEQ ID NO: 21085.	
XX	
KW Protein identification; signal transduction pathway; metabolic pathway;	
KW hybridisation assay; genetic mapping; gene expression control; promoter;	
KW termination sequence.	
XX	
OS Arabidopsis thaliana.	
XX	
PN EP1033405-A2.	
XX	
PD 06-SEP-2000.	
XX	
PF 25-FEB-2000; 2000EP-0301439.	
XX	
PR 25-FEB-1999;	99US-0121825.
PR 05-MAR-1999;	99US-0123180.
PR 09-MAR-1999;	99US-0123548.
PR 23-MAR-1999;	99US-0125788.
PR 25-MAR-1999;	99US-0126264.
PR 29-MAR-1999;	99US-0126785.
PR 01-APR-1999;	99US-0127462.
PR 06-APR-1999;	99US-0128234.
PR 08-APR-1999;	99US-0128714.
PR 16-APR-1999;	99US-0129845.
PR 19-APR-1999;	99US-0130077.
PR 21-APR-1999;	99US-0130449.
PR 23-APR-1999;	99US-0130510.
PR 23-APR-1999;	99US-0130891.
PR 28-APR-1999;	99US-0131449.
PR 30-APR-1999;	99US-0132048.
PR 30-APR-1999;	99US-0132407.

PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139730.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 21-JUL-1999; 99US-0145089.

PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145219.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.

PR	21-OCT-1999;	99US-0160768.	PR	14-MAY-1999;	99US-0134219.
PR	21-OCT-1999;	99US-0160770.	PR	14-MAY-1999;	99US-0134221.
PR	21-OCT-1999;	99US-0160814.	PR	14-MAY-1999;	99US-0134370.
PR	21-OCT-1999;	99US-0160815.	PR	18-MAY-1999;	99US-0134768.
PR	22-OCT-1999;	99US-0160980.	PR	19-MAY-1999;	99US-0134941.
PR	22-OCT-1999;	99US-0160981.	PR	20-MAY-1999;	99US-0135124.
PR	22-OCT-1999;	99US-0160989.	PR	21-MAY-1999;	99US-0135353.
PR	25-OCT-1999;	99US-0161404.	PR	24-MAY-1999;	99US-0135629.
PR	25-OCT-1999;	99US-0161405.	PR	25-MAY-1999;	99US-0136021.
PR	25-OCT-1999;	99US-0161406.	PR	27-MAY-1999;	99US-0136392.
PR	26-OCT-1999;	99US-0161359.	PR	28-MAY-1999;	99US-0136782.
PR	26-OCT-1999;	99US-0161360.	PR	01-JUN-1999;	99US-0137222.
PR	26-OCT-1999;	99US-0161361.	PR	03-JUN-1999;	99US-0137528.
PR	28-OCT-1999;	99US-0161920.	PR	04-JUN-1999;	99US-0137502.
PR	28-OCT-1999;	99US-0161992.	PR	07-JUN-1999;	99US-0137724.
PR	28-OCT-1999;	99US-0161993.	PR	08-JUN-1999;	99US-0138094.
PR	29-OCT-1999;	99US-0162142.	PR	10-JUN-1999;	99US-0138540.
Query Match 3.7%; Score 8; DB 21; Length 82;			PR	10-JUN-1999;	99US-0138847.
Best Local Similarity 100.0%; Pred. No. 7.4;			PR	14-JUN-1999;	99US-0139119.
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			PR	16-JUN-1999;	99US-0139452.
OY	120 AAAAPAAP 127		PR	16-JUN-1999;	99US-0139453.
Db	8 AAAAPAAP 15		PR	17-JUN-1999;	99US-0139453.
			PR	18-JUN-1999;	99US-0139454.
RESULT 12			PR	18-JUN-1999;	99US-0139455.
AAG19326			PR	18-JUN-1999;	99US-0139456.
ID AAG19326 standard; Protein; 94 AA.			PR	18-JUN-1999;	99US-0139457.
AC AAG19326;			PR	18-JUN-1999;	99US-0139458.
XX			PR	18-JUN-1999;	99US-0139459.
XX	17-OCT-2000 (first entry)		PR	18-JUN-1999;	99US-0139460.
XX	Arabidopsis thaliana protein fragment SEQ ID NO: 21084.		PR	18-JUN-1999;	99US-0139461.
XX	Protein identification; signal transduction pathway; metabolic pathway;		PR	18-JUN-1999;	99US-0139462.
KW	hybridisation assay; genetic mapping; gene expression control; promoter;		PR	18-JUN-1999;	99US-0139463.
KW	termination sequence.		PR	18-JUN-1999;	99US-0139750.
OS	Arabidopsis thaliana.		PR	18-JUN-1999;	99US-0139763.
XX			PR	21-JUN-1999;	99US-0139817.
XX	EPI033405-A2.		PR	22-JUN-1999;	99US-0139899.
XX	06-SEP-2000.		PR	23-JUN-1999;	99US-0140353.
XX	25-FEB-2000; 2000EP-0301439.		PR	23-JUN-1999;	99US-0140354.
PR	25-FEB-1999;	99US-0121825.	PR	24-JUN-1999;	99US-0140695.
PR	05-MAR-1999;	99US-0123180.	PR	28-JUN-1999;	99US-0140823.
PR	09-MAR-1999;	99US-0123548.	PR	30-JUN-1999;	99US-0140991.
PR	23-MAR-1999;	99US-0125788.	PR	01-JUL-1999;	99US-0141287.
PR	25-MAR-1999;	99US-0126264.	PR	02-JUL-1999;	99US-0142154.
PR	29-MAR-1999;	99US-0126785.	PR	06-JUL-1999;	99US-0142055.
PR	01-APR-1999;	99US-0127462.	PR	08-JUL-1999;	99US-0142390.
PR	06-APR-1999;	99US-0128234.	PR	09-JUL-1999;	99US-0142803.
PR	08-APR-1999;	99US-0128714.	PR	12-JUL-1999;	99US-0142920.
PR	16-APR-1999;	99US-0129845.	PR	13-JUL-1999;	99US-0142977.
PR	19-APR-1999;	99US-0130077.	PR	14-JUL-1999;	99US-0143542.
PR	21-APR-1999;	99US-0130449.	PR	15-JUL-1999;	99US-0143624.
PR	23-APR-1999;	99US-0130510.	PR	16-JUL-1999;	99US-0144005.
PR	23-APR-1999;	99US-0130891.	PR	16-JUL-1999;	99US-0144085.
PR	28-APR-1999;	99US-0131449.	PR	19-JUL-1999;	99US-0144086.
PR	30-APR-1999;	99US-0132048.	PR	19-JUL-1999;	99US-0144325.
PR	30-APR-1999;	99US-0132407.	PR	19-JUL-1999;	99US-0144331.
PR	04-MAY-1999;	99US-0132484.	PR	19-JUL-1999;	99US-0144332.
PR	05-MAY-1999;	99US-0132485.	PR	19-JUL-1999;	99US-0144333.
PR	06-MAY-1999;	99US-0132486.	PR	20-JUL-1999;	99US-0144335.
PR	06-MAY-1999;	99US-0132487.	PR	20-JUL-1999;	99US-0144632.
PR	07-MAY-1999;	99US-0132863.	PR	20-JUL-1999;	99US-0144884.
PR	11-MAY-1999;	99US-0134256.	PR	21-JUL-1999;	99US-0144884.
PR	14-MAY-1999;	99US-0134256.	PR	21-JUL-1999;	99US-0145086.
			PR	21-JUL-1999;	99US-0145088.
			PR	22-JUL-1999;	99US-0145085.
			PR	22-JUL-1999;	99US-0145087.
			PR	22-JUL-1999;	99US-0145089.
			PR	22-JUL-1999;	99US-0145192.
			PR	23-JUL-1999;	99US-0145145.
			PR	23-JUL-1999;	99US-0145218.
			PR	23-JUL-1999;	99US-0145224.

PR	26-JUL-1999;	99US-0145276.
PR	27-JUL-1999;	99US-0145913.
PR	27-JUL-1999;	99US-0145918.
PR	27-JUL-1999;	99US-0145919.
PR	28-JUL-1999;	99US-0145951.
PR	02-AUG-1999;	99US-0146386.
PR	02-AUG-1999;	99US-0146388.
PR	02-AUG-1999;	99US-0146389.
PR	03-AUG-1999;	99US-0147038.
PR	04-AUG-1999;	99US-0147204.
PR	04-AUG-1999;	99US-0147302.
PR	05-AUG-1999;	99US-0147302.
PR	05-AUG-1999;	99US-0147192.
PR	05-AUG-1999;	99US-0147260.
PR	06-AUG-1999;	99US-0147303.
PR	06-AUG-1999;	99US-0147303.
PR	06-AUG-1999;	99US-0147416.
PR	09-AUG-1999;	99US-0147493.
PR	09-AUG-1999;	99US-0147935.
PR	10-AUG-1999;	99US-0148171.
PR	11-AUG-1999;	99US-0148319.
PR	12-AUG-1999;	99US-0148341.
PR	13-AUG-1999;	99US-0148565.
PR	13-AUG-1999;	99US-0148684.
PR	16-AUG-1999;	99US-0149368.
PR	17-AUG-1999;	99US-0149175.
PR	18-AUG-1999;	99US-0149426.
PR	20-AUG-1999;	99US-0149722.
PR	20-AUG-1999;	99US-0149723.
PR	20-AUG-1999;	99US-0149929.
PR	23-AUG-1999;	99US-0149902.
PR	23-AUG-1999;	99US-0149930.
PR	25-AUG-1999;	99US-0150566.
PR	26-AUG-1999;	99US-0150884.
PR	27-AUG-1999;	99US-0151065.
PR	27-AUG-1999;	99US-0151066.
PR	27-AUG-1999;	99US-0151080.
PR	30-AUG-1999;	99US-0151303.
PR	31-AUG-1999;	99US-0151438.
PR	01-SEP-1999;	99US-0151930.
PR	07-SEP-1999;	99US-0152363.
PR	10-SEP-1999;	99US-0153070.
PR	13-SEP-1999;	99US-0153758.
PR	15-SEP-1999;	99US-0154018.
PR	16-SEP-1999;	99US-0154039.
PR	20-SEP-1999;	99US-0154779.
PR	22-SEP-1999;	99US-0155139.
PR	23-SEP-1999;	99US-0155480.
PR	24-SEP-1999;	99US-0155653.
PR	28-SEP-1999;	99US-0156458.
PR	29-SEP-1999;	99US-0156596.
PR	04-OCT-1999;	99US-0157117.
PR	05-OCT-1999;	99US-0157753.
PR	06-OCT-1999;	99US-0157863.
PR	07-OCT-1999;	99US-0158023.
PR	08-OCT-1999;	99US-0158232.
PR	12-OCT-1999;	99US-0158369.
PR	13-OCT-1999;	99US-0159293.
PR	13-OCT-1999;	99US-0159294.
PR	13-OCT-1999;	99US-0159299.
PR	14-OCT-1999;	99US-0159325.
PR	14-OCT-1999;	99US-0159330.
PR	14-OCT-1999;	99US-0159331.
PR	14-OCT-1999;	99US-0159637.
PR	14-OCT-1999;	99US-0159638.
PR	18-OCT-1999;	99US-0159584.
PR	21-OCT-1999;	99US-0160741.
PR	21-OCT-1999;	99US-0160767.
PR	21-OCT-1999;	99US-0160768.
PR	21-OCT-1999;	99US-0160770.
PR	21-OCT-1999;	99US-0160814.
PR	21-OCT-1999;	99US-0160815.
PR	22-OCT-1999;	99US-0160980.
PR	22-OCT-1999;	99US-0160981.
PR	22-OCT-1999;	99US-0160989.
PR	26-OCT-1999;	99US-0161404.
PR	25-OCT-1999;	99US-0161405.
PR	25-OCT-1999;	99US-0161406.
PR	26-OCT-1999;	99US-0161359.
PR	26-OCT-1999;	99US-0161360.
PR	26-OCT-1999;	99US-0161361.
PR	28-OCT-1999;	99US-0161920.
PR	28-OCT-1999;	99US-0161992.
PR	28-OCT-1999;	99US-0161993.
PR	29-OCT-1999;	99US-0162142.
Query Match 3.7%; Score 8; DB 21; Length 94;		
Best Local Similarity 100.0%; Pred. No. 8.2;		
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
QY	120 AAAAPAAP 127	
Db	20 AAAAPAAP 27	
RESULT 13		
AAG19325		
ID	AAG19325 standard; Protein; 106 AA.	
XX		
AC	AAG19325;	
XX		
DT	17-OCT-2000 (first entry)	
XX		
XX	Arabidopsis thaliana protein fragment SEQ ID NO: 21083.	
XX	Protein identification; signal transduction pathway; metabolic pathway;	
KW	hybridisation assay; genetic mapping; gene expression control; promoter;	
KW	termination sequence.	
XX		
OS	Arabidopsis thaliana.	
XX		
PN	EP1033405-A2.	
XX		
PD	06-SEP-2000.	
XX		
PF	25-FEB-2000; 2000EP-0301439.	
XX		
PR	25-FEB-1999;	99US-0121825.
PR	05-MAR-1999;	99US-0123180.
PR	09-MAR-1999;	99US-0123548.
PR	23-MAR-1999;	99US-0125788.
PR	25-MAR-1999;	99US-0126264.
PR	29-MAR-1999;	99US-0126785.
PR	01-APR-1999;	99US-0127462.
PR	06-APR-1999;	99US-0128234.
PR	08-APR-1999;	99US-0128714.
PR	16-APR-1999;	99US-0129845.
PR	19-APR-1999;	99US-0130077.
PR	21-APR-1999;	99US-0130449.
PR	23-APR-1999;	99US-0130510.
PR	23-APR-1999;	99US-0130891.
PR	28-APR-1999;	99US-0131449.
PR	30-APR-1999;	99US-0132048.
PR	30-APR-1999;	99US-0132407.
PR	04-MAY-1999;	99US-0132484.
PR	05-MAY-1999;	99US-0132485.
PR	06-MAY-1999;	99US-0132486.
PR	06-MAY-1999;	99US-0132487.
PR	07-MAY-1999;	99US-0132863.
PR	11-MAY-1999;	99US-0134256.
PR	14-MAY-1999;	99US-0134218.
PR	14-MAY-1999;	99US-0134219.
PR	14-MAY-1999;	99US-0134221.
PR	14-MAY-1999;	99US-0134370.
PR	18-MAY-1999;	99US-0134768.
PR	19-MAY-1999;	99US-0134941.
PR	20-MAY-1999;	99US-0135124.
PR	21-MAY-1999;	99US-0135353.

PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139432.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 24-JUN-1999; 99US-0140823.
PR 28-JUN-1999; 99US-0140981.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.

```
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match          3.7%; Score 8; DB 21; Length 106;
Best Local Similarity 100.0%; Pred. No. 9.1;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 120 AAAAPAAP 127
Db 32 AAAAPAAP 39
|||||

RESULT 14
ID AAB40893 standard; Protein: 121 AA.
AC AAB40893;
XX
XX 08-FEB-2001 (first entry)
DT
DE Human ORFX ORF657 polypeptide sequence SEQ ID NO:1314.
DE
XX
XX Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW vulnery; antiposiatric; antiparkinsonian; nootropic; neuroprotective;
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antinflammatory;
KW antiviral; antibacterial; antifungal; antirheumatic; antihyroid;
KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW thrombosis; contraceptive.
XX
XX Homo sapiens.
OS
XX WO200058473-A2.
XX
XX 05-OCT-2000.
XX
XX 31-MAR-2000; 2000WO-US08621.
XX
XX 31-MAR-1999; 99US-0127607.
XX 02-APR-1999; 99US-0127636.
XX 05-APR-1999; 99US-0127728.
XX 30-MAR-2000; 2000US-0540763.
XX
XX (CURA-) CURAGEN CORP.
PA
XX Shinkets RA, Leach M;
XX
XX WPI; 2000-602362/57.
XX N-PSDB; AAC75102.
XX
XX Novel nucleic acids and peptides derived from open reading frame X,
XX useful for treating e.g. cancers, proliferative disorders,
XX neurodegenerative disorders and cardiovascular disease -
XX
XX Claim 11; Page 1131; 5507pp; English.
XX
XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
XX which represent the human ORFX open reading frames 1 to 3161. The ORFX
XX sequences have activities such as: cytostatic; hepatotropic; vulnerary;
XX antiposiatric; antiparkinsonian; nootropic; neuroprotective;
XX osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
XX immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
XX antidiabetic; hypotensive; dermatological; immunosuppressive;
XX antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
XX and thymoid; and antianaemic. The sequences can be used for determining
```


PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 23-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 22-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 14-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.

PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 3.7%; Score 8; DB 21; Length 124;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 130 AAEPA AAA 137
| | | | |
Db 19 AAEPA AAA 26

Search completed: February 12, 2003, 13:57:51
Job time : 38 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2003, 13:57:05 ; Search time 20 seconds
(without alignments)
322.181 Million cell updates/sec

Title: US-09-923-236-2
Perfect score: 219
Sequence: 1 MKLLWACIVCAFARRRFE.....TAKPAAPPEPHSPSLEQANQ 219

Scoring table: OLIGO

Searched: 262574 seqs, 29422922 residues

Word size : 0

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued Patents_AA.*

- 1: /cgn2_6/ptodata/1/iaa/5A_COMB.pap.*
- 2: /cgn2_6/ptodata/1/iaa/5B_COMB.pap.*
- 3: /cgn2_6/ptodata/1/iaa/6A_COMB.pap.*
- 4: /cgn2_6/ptodata/1/iaa/6B_COMB.pap.*
- 5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pap.*
- 6: /cgn2_6/ptodata/1/iaa/backfiles1.pap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	219	100.0	219	4	US-09-527-345-2
2	9	4.1	322	2	US-08-428-414A-2
3	8	3.7	383	4	US-09-471-396-3
4	8	3.7	412	4	US-09-471-396-1
5	8	3.7	575	4	US-08-653-648A-5
6	8	3.7	4545	2	US-08-804-227C-14
7	8	3.7	4550	2	US-08-804-227C-8
8	8	3.7	4550	2	US-08-804-198-2
9	7	3.2	31	4	US-09-248-588-27
10	7	3.2	76	6	5273901-11
11	7	3.2	76	6	5482709-10
12	7	3.2	76	6	5496550-6
13	7	3.2	118	3	US-08-301-162-10
14	7	3.2	118	4	US-09-461-920-10
15	7	3.2	118	4	US-09-968-927-10
16	7	3.2	156	1	US-08-074-121-5
17	7	3.2	156	5	PCT-US94-06447-5
18	7	3.2	159	3	US-08-481-435-13
19	7	3.2	180	6	5273901-7
20	7	3.2	180	6	5482709-6
21	7	3.2	283	4	US-09-135-021-6
22	7	3.2	315	3	US-09-362-318-2
23	7	3.2	335	2	US-08-405-175A-6
24	7	3.2	401	3	US-08-289-222E-3
25	7	3.2	401	4	US-09-054-526B-3
26	7	3.2	405	1	US-08-370-193A-9
27	7	3.2	415	4	US-08-927-219-8

28	7	3.2	430	4	US-09-105-058C-21	Sequence 21, Appl
29	7	3.2	501	2	US-08-288-508C-2	Sequence 2, Appli
30	7	3.2	532	3	US-08-481-435-12	Sequence 12, Appl
31	7	3.2	553	3	US-08-481-435-11	Sequence 11, Appl
32	7	3.2	587	1	US-08-398-008A-2	Sequence 2, Appli
33	7	3.2	587	2	US-08-893-333-2	Sequence 2, Appli
34	7	3.2	676	4	US-09-135-021-2	Sequence 2, Appli
35	7	3.2	676	4	US-09-135-020-2	Sequence 2, Appli
36	7	3.2	676	4	US-09-135-010A-2	Sequence 2, Appli
37	7	3.2	676	4	US-09-444-871-2	Sequence 2, Appli
38	7	3.2	676	4	US-09-634-920-2	Sequence 2, Appli
39	7	3.2	676	4	US-09-597-735-2	Sequence 2, Appli
40	7	3.2	676	4	US-09-444-295-2	Sequence 2, Appli
41	7	3.2	676	4	US-09-597-732-2	Sequence 2, Appli
42	7	3.2	677	4	US-09-177-650-4	Sequence 4, Appli
43	7	3.2	823	3	US-08-481-435-4	Sequence 4, Appli
44	7	3.2	836	3	US-08-481-435-9	Sequence 9, Appli
45	7	3.2	844	3	US-08-481-435-7	Sequence 7, Appli

ALIGNMENTS

RESULT 1
US-09-527-345-2
; Sequence 2, Application US/09527345
; Patent No. 6331413
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; APPLICANT: Adler, David A.
; TITLE OF INVENTION: SECRETED SALIVARY ZSIG63 POLYPEPTIDE
; FILE REFERENCE: 97-71
; CURRENT APPLICATION NUMBER: US/09/527,345
; CURRENT FILING DATE: 1999-03-17
; PRIOR APPLICATION NUMBER: US 60/124,820
; PRIOR FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 219
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-527-345-2

Query Match	100.0%;	Score 219;	DB 4;	Length 219;			
Best Local Similarity	100.0%;	Pred. No. 8.3e-204;					
Matches 219;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;			
QY	1	MKLLWACIVCAFARRRPF	IGEDDNDGHP	PLHPSLNIPY	GIRNLPPLXYR	PVNTVP	60
Db	1	MKLLWACIVCAFARRRPF	IGEDDNDGHP	PLHPSLNIPY	GIRNLPPLXYR	PVNTVP	60
QY	61	SYPGNTYDTGLSPY	WILLTSPGFPY	VYHIRGF	PLATQLNVP	PLPRGFP	VPSPRFFSA 120
Db	61	SYPGNTYDTGLSPY	WILLTSPGFPY	VYHIRGF	PLATQLNVP	PLPRGFP	VPSPRFFSA 120
QY	121	AAAPAAPIAAEPAAP	LAATPVAAEPA	ACAPVAAEPA	APVAAEPA	APVAAEPA	180
Db	121	AAAPAAPIAAEPAAP	LAATPVAAEPA	ACAPVAAEPA	APVAAEPA	APVAAEPA	180
QY	181	EAPYGVFAAEPSFA	EPATAPAKPA	PEPHSPSLEQ	ANQ 219		
Db	181	EAPYGVFAAEPSFA	EPATAPAKPA	PEPHSPSLEQ	ANQ 219		
RESULT 2							
US-08-428-414A-2							
; Sequence 2, Application US/08428414A							
; Patent No. 5912166							
; GENERAL INFORMATION:							
; APPLICANT: Reed, Steven G.							
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF							
; TITLE OF INVENTION: LEISHMANIASIS							

NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/428,414A
FILING DATE: 21-APR-1995
CLASSIFICATION: 436
ATTORNEY/AGENT INFORMATION:
NAME: Kadlecek, Ann T.
REGISTRATION NUMBER: 39,244
REFERENCE/DOCKET NUMBER: 210121.407
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
TELEX: 3723836 SEEDANDBERRY
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 322 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-428-414A-2

Query Match 4.1%; Score 9; DB 2; Length 322;
Best Local Similarity 100.0%; Pred. No. 0.57; Indels 0; Gaps 0;
Matches 9; Conservative 0; Mismatches 0;

QY 130 AAEPAAAAAP 138
|||||
DB 289 AAEPAAAAAP 297

RESULT 3
US-09-471-396-3
Sequence 3, Application US/09471396
Patent No. 6458359
GENERAL INFORMATION:
APPLICANT: BEDATE, Carlos Alonso
APPLICANT: REQUENA ROLANIA, Jose Maria
APPLICANT: SOTO ALVAREZ, Manuel
TITLE OF INVENTION: CHIMERIC GENE FORMED OF THE DNA SEQUENCES THAT ENCODE
TITLE OF INVENTION: THE ANTIGENIC DETERMINANTS OF FOUR PROTEINS OF
TITLE OF INVENTION: L-INFANTUM AND PROTEIN ENCODED BY SAIDI GENE, AND ...
FILE REFERENCE: bedateza.seq
CURRENT APPLICATION NUMBER: US/09/471,396
CURRENT FILING DATE: 1999-12-23
PRIOR APPLICATION NUMBER: 60/113,825
PRIOR FILING DATE: 1998-12-23
NUMBER OF SEQ ID NOS: 12
SOFTWARE: PatentIn ver. 2.1
SEQ ID NO 3
LENGTH: 383
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Protein Q
US-09-471-396-3

Query Match 3.7%; Score 8; DB 4; Length 383;
Best Local Similarity 100.0%; Pred. No. 6.2; Indels 0; Gaps 0;
Matches 8; Conservative 0; Mismatches 0;

QY 120 AAAAPAAP 127

Db 354 AAAAPAAP 361
|||||
RESULT 4
US-09-471-396-1
Sequence 1, Application US/09471396
Patent No. 6458359
GENERAL INFORMATION:
APPLICANT: BEDATE, Carlos Alonso
APPLICANT: REQUENA ROLANIA, Jose Maria
APPLICANT: SOTO ALVAREZ, Manuel
TITLE OF INVENTION: CHIMERIC GENE FORMED OF THE DNA SEQUENCES THAT ENCODE
TITLE OF INVENTION: THE ANTIGENIC DETERMINANTS OF FOUR PROTEINS OF
TITLE OF INVENTION: L-INFANTUM AND PROTEIN ENCODED BY SAIDI GENE, AND ...
FILE REFERENCE: bedateza.seq
CURRENT APPLICATION NUMBER: US/09/471,396
CURRENT FILING DATE: 1999-12-23
PRIOR APPLICATION NUMBER: 60/113,825
PRIOR FILING DATE: 1998-12-23
NUMBER OF SEQ ID NOS: 12
SOFTWARE: PatentIn ver. 2.1
SEQ ID NO 1
LENGTH: 412
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Protein Q
US-09-471-396-1

Query Match 3.7%; Score 8; DB 4; Length 412;
Best Local Similarity 100.0%; Pred. No. 6.6; Indels 0; Gaps 0;
Matches 8; Conservative 0; Mismatches 0;

QY 120 AAAAPAAP 127
|||||
DB 383 AAAAPAAP 390

RESULT 5
US-08-653-648A-5
Sequence 5, Application US/08653648A
Patent No. 6379945
GENERAL INFORMATION:
APPLICANT: Jepson, Ian
APPLICANT: Greenland, Andrew
APPLICANT: Martinez, Alberto
TITLE OF INVENTION: A Gene Switch
FILE REFERENCE: PPD50047/US
CURRENT APPLICATION NUMBER: US/08/653,648A
CURRENT FILING DATE: 1996-05-24
PRIOR APPLICATION NUMBER: GB 9510759.5
PRIOR FILING DATE: 1995-05-26
PRIOR APPLICATION NUMBER: GB 9605656.9
PRIOR FILING DATE: 1996-03-18
PRIOR APPLICATION NUMBER: GB 9513882.2
PRIOR FILING DATE: 1995-07-07
PRIOR APPLICATION NUMBER: GB 9517316.7
PRIOR FILING DATE: 1995-08-24
NUMBER OF SEQ ID NOS: 65
SOFTWARE: PatentIn version 3.0
SEQ ID NO 5
LENGTH: 575
TYPE: PRT
ORGANISM: Heliothis virescens
US-08-653-648A-5

Query Match 3.7%; Score 8; DB 4; Length 575;
Best Local Similarity 100.0%; Pred. No. 9; Indels 0; Gaps 0;
Matches 8; Conservative 0; Mismatches 0;

QY 140 TATPVAAP 147
|||||

Db 552 TATPVAE 559

RESULT 6

US-08-804-227C-14
; Sequence 14, Application US/08804227C
; Patent No. 5876991
; GENERAL INFORMATION:
; APPLICANT: DeHoff, Bradley S.
; APPLICANT: Kilstoss, Stuart A.
; APPLICANT: Rostock, Paul R., Jr.
; APPLICANT: Sutton, Kimberly L.
; TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: THOMAS G. PLANT 1501
; STREET: LILLY CORPORATE CENTER
; CITY: INDIANAPOLIS
; STATE: IN
; COUNTRY: USA
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: ASCII(DOS) Text only
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/804.227C
; FILING DATE: February 21, 1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Plant, Thomas, G.
; REGISTRATION NUMBER: 35,784
; REFERENCE/DOCKET NUMBER: X-8231
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-276-2459
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4545 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-804-227C-14

Query Match 3.7%; Score 8; DB 2; Length 4545;
Best Local Similarity 100.0%; Pred. No. 59;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 131 AEPAAAP 138

Db 2620 AEPAAAP 2627

RESULT 7

US-08-804-227C-8
; Sequence 8, Application US/08804227C
; Patent No. 5876991
; GENERAL INFORMATION:
; APPLICANT: DeHoff, Bradley S.
; APPLICANT: Kilstoss, Stuart A.
; APPLICANT: Rostock, Paul R., Jr.
; APPLICANT: Sutton, Kimberly L.
; TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: THOMAS G. PLANT 1501
; STREET: LILLY CORPORATE CENTER
; CITY: INDIANAPOLIS
; STATE: IN
; COUNTRY: USA
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: ASCII(DOS) Text only
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/804.227C
; FILING DATE: February 21, 1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Plant, Thomas, G.
; REGISTRATION NUMBER: 35,784
; REFERENCE/DOCKET NUMBER: X-8231
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-276-2459
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4550 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
US-08-804-227C-8

Query Match 3.7%; Score 8; DB 2; Length 4550;
Best Local Similarity 100.0%; Pred. No. 59;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 131 AEPAAAP 138

Db 2625 AEPAAAP 2632

RESULT 8

US-08-804-198-2
; Sequence 2, Application US/08804198
; Patent No. 5945320
; GENERAL INFORMATION:
; APPLICANT: Burgett, Stanley G.
; APPLICANT: Kilstoss, Stuart A.
; APPLICANT: Rao, Nagaraja R.
; APPLICANT: Richardson, Mark A.
; APPLICANT: Rostock, Paul R., Jr.
; TITLE OF INVENTION: PLATENOLIDE SYNTHASE GENE
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PAUL R. CANTRELL 1138
; STREET: LILLY CORPORATE CENTER
; CITY: INDIANAPOLIS
; STATE: IN
; COUNTRY: USA
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Macintosh
; OPERATING SYSTEM: Macintosh 7.0
; SOFTWARE: Microsoft Word 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/804.198
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: CANTRELL, PAUL R.
; REGISTRATION NUMBER: 36,470
; REFERENCE/DOCKET NUMBER: P9113
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-276-3885
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4550 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
US-08-804-198-2

Query Match 3.7%; Score 8; DB 2; Length 4550;

Best Local Similarity 100.0%; Pred. No. 59;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 131 AEPAAAP 138

Db 2625 AEPAAAP 2632

RESULT 9

US-09-248-588-27
; Sequence 27, Application US/09248588
; Patent No. 6231864
; GENERAL INFORMATION:
; APPLICANT: Birkett, Ashley J.
; TITLE OF INVENTION: Strategically Modified Hepatitis B Core Proteins and
; FILE REFERENCE: SYN-101 4564/69529
; CURRENT APPLICATION NUMBER: US/09/248,588
; CURRENT FILING DATE: 1999-02-11
; EARLIER APPLICATION NUMBER: 60/074537
; EARLIER FILING DATE: 1998-02-12
; NUMBER OF SEQ ID NOS: 113
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 27
; LENGTH: 31
; TYPE: PRT
; ORGANISM: Cryptosporidium parvum
; PUBLICATION INFORMATION:
; PATENT DOCUMENT NUMBER: WO 98/07320
US-09-248-588-27

Query Match 3.2%; Score 7; DB 4; Length 31;
Best Local Similarity 100.0%; Pred. No. 5.8;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 155 AEPAAE 161

Db 14 AEPAAE 20

RESULT 10

5273901-11
; Patent No. 5273901
; APPLICANT: JACOBSON, JAMES W.; STRAUSBERG, ROBERT L.; WILSON,
; SUSAN D.; POPE, SHARON H.; STRAUSBERG, SUSAN L.; RUFF, MICHAEL D.;
; AUGUSTINE, PATRICIA C.; DANFORTH, HARRY D.
; TITLE OF INVENTION: GENETICALLY ENGINEERED COCCIDIOSIS
; SPOOROZITE 21.5 KB ANTIGEN, AC-6B
; NUMBER OF SEQUENCES: 11
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/581,693
; FILING DATE: 12-SEP-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 215,162
; FILING DATE: 05-JUL-1988
; APPLICATION NUMBER: 746,520
; FILING DATE: 19-JUN-1985
; APPLICATION NUMBER: 627,811
; FILING DATE: 05-JUL-1984
; SEQ ID NO:11:
; LENGTH:76
5273901-11

Query Match 3.2%; Score 7; DB 6; Length 76;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 120 AAAAPAA 126

Db 20 AAAAPAA 26

RESULT 11

5482709-10
; Patent No. 5482709
; APPLICANT: JACOBSON, JAMES W.; STRAUSBERG, ROBERT L.; WILSON,
; SUSAN D.; POPE, SHARON H.; STRAUSBERG, SUSAN L.; RUFF, MICHAEL D.;
; AUGUSTINE, PATRICIA C.; DANFORTH, HARRY D.
; TITLE OF INVENTION: EIMERIA ANTIGENIC COMPOSITION WHICH
; ELICITS ANTIBODIES AGAINST AVIAN COCCIDIOSIS
; NUMBER OF SEQUENCES: 10
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/148,432
; FILING DATE: 08-NOV-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 581,693
; FILING DATE: 12-SEP-1990
; APPLICATION NUMBER: 215,162
; FILING DATE: 05-JUL-1989
; APPLICATION NUMBER: 746,520
; FILING DATE: 19-JUN-1985
; APPLICATION NUMBER: 627,811
; FILING DATE: 05-JUL-1984
; SEQ ID NO:10:
; LENGTH: 76
5482709-10

Query Match 3.2%; Score 7; DB 6; Length 76;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 120 AAAAPAA 126

Db 20 AAAAPAA 26

RESULT 12

5496550-6
; Patent No. 5496550
; APPLICANT: WALLACH, MICHAEL; PUGATSCH, THEA; MENCHER, DAVID
; TITLE OF INVENTION: METHOD OF REDUCING THE OUTPUT OF EIMERIA
; OOCYSTS FROM A NEWBORN CHICK
; NUMBER OF SEQUENCES: 10
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/108,763
; FILING DATE: 17-AUG-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 642,219
; FILING DATE: 16-JAN-1991
; APPLICATION NUMBER: 310,603
; FILING DATE: 14-FEB-1989
; APPLICATION NUMBER: 155,245
; FILING DATE: 12-FEB-1988
; APPLICATION NUMBER: 896,611
; FILING DATE: 14-AUG-1986
; SEQ ID NO:6:
; LENGTH: 76
5496550-6

Query Match 3.2%; Score 7; DB 6; Length 76;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 120 AAAAPAA 126

Db 24 AAAAPAA 30

RESULT 13

US-08-301-162-10
; Sequence 10, Application US/08301162
; Patent No. 6022546
; GENERAL INFORMATION:
; APPLICANT: Knapp, Stefan
; APPLICANT: Ziegelmaier, Robert
; APPLICANT: Kupper, Hans

;
; TITLE OF INVENTION: Toxoplasma Gondii Antigens, The
; TITLE OF INVENTION: Preparation Thereof and the Use Thereof
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; STREET: 1300 I Street, N.W., Suite 700
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/301,162
; FILING DATE: 06-DEC-1990
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US/08/167,128
; FILING DATE: 06-DEC-1990
; APPLICATION NUMBER: US 07/623,086
; FILING DATE: 06-DEC-1990
; APPLICATION NUMBER: DE P3940598.2
; FILING DATE: 08-DEC-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Fleshner, Raz E.
; REGISTRATION NUMBER: 34,331
; REFERENCE/DOCKET NUMBER: 02481.1005-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-408-4000
; TELEFAX: 202-408-4400
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 118 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-301-162-10

Query Match 3.2%; Score 7; DB 3; Length 118;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 120 AAAAPAA 126
| | | | |
Db 84 AAAAPAA 90

RESULT 14
US-09-461-240-10
; Sequence 10, Application US/09461240
; Patent No. 6326008
; GENERAL INFORMATION:
; APPLICANT: Knapp, Stefan
; Ziegelmaier, Robert
; Kupper, Hans
; TITLE OF INVENTION: Toxoplasma Gondii Antigens, The
; Preparation Thereof and the Use Thereof
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; Danner
; STREET: 1300 I Street, N.W., Suite 700
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/461,240
; FILING DATE: 16-Dec-1999
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/301,162
; FILING DATE: <Unknown>
; APPLICATION NUMBER: US/08/167,128
; FILING DATE: <Unknown>
; APPLICATION NUMBER: US 07/623,086
; FILING DATE: 06-DEC-1990
; APPLICATION NUMBER: DE P3940598.2
; FILING DATE: 08-DEC-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Fleshner, Raz E.
; REGISTRATION NUMBER: 34,331
; REFERENCE/DOCKET NUMBER: 02481.1005-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-408-4000
; TELEFAX: 202-408-4400
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 118 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 10:
US-09-461-240-10

Query Match 3.2%; Score 7; DB 4; Length 118;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 120 AAAAPAA 126
| | | | |
Db 84 AAAAPAA 90

RESULT 15
US-09-968-927-10
; Sequence 10, Application US/09968927
; Patent No. 6419925
; GENERAL INFORMATION:
; APPLICANT: Knapp, Stefan
; Ziegelmaier, Robert
; Kupper, Hans
; TITLE OF INVENTION: Toxoplasma Gondii Antigens, The
; Preparation Thereof and the Use Thereof
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; Danner
; STREET: 1300 I Street, N.W., Suite 700
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/968,927
; FILING DATE: 03-Oct-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/301,162
; FILING DATE: <Unknown>
; APPLICATION NUMBER: US/08/167,128
; FILING DATE: <Unknown>

```

; APPLICATION NUMBER: US 07/623,086
; FILING DATE: 06-DEC-1990
; APPLICATION NUMBER: DE P3940598.2
; FILING DATE: 08-DEC-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Fleshner, Raz E.
; REGISTRATION NUMBER: 34,331
; REFERENCE/DOCKET NUMBER: 02481.1005-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-408-4000
; TELEFAX: 202-408-4400
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 118 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 10:
US-09-968-927-10
Query Match          3.2%  Score 7; DB 4; Length 118;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 120 AAAPAA 126
Db 84 AAAPAA 90

Search completed: February 12, 2003, 14:00:11
Job time : 23 secs
```


GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2003, 13:59:10 ; Search time 14 Seconds
(without alignments)
399.657 Million cell updates/sec

Title: US-09-923-236-2

perfect score: 219
Sequence: 1 MKLLWACIVCVAFARKRF.....TAKPAAPEHPSPSLEQANQ 219

Scoring table: OLIGO

scoring cable. 01150
Gapop 60.0 , Gapext 60.0

Searched: 140259 seqs, 25548876 residues

Word size : 0

Total number of hits satisfying chosen parameters: 140259

Minimum DB seq length: 0

Maximum DB seq length: 2000000000
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :	Published_Applications_RA:*
1:	/cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pcp:*
2:	/cgn2_6/ptodata/1/pubpaa/PTCT_NEW_PUB.pcp:*
3:	/cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pcp:*
4:	/cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pcp:*
5:	/cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pcp:*
6:	/cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pcp:*
7:	/cgn2_6/ptodata/1/pubpaa/PTCTUS_PUBCOMB.pcp:*
8:	/cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pcp:*
9:	/cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pcp:*
10:	/cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pcp:*
11:	/cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pcp:*
12:	/cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pcp:*
13:	/cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pcp:*
14:	/cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pcp:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query			ID	Description
	Score	Match	Length		
1	219	100.0	219	US-09-922-469-2	Sequence 2, Appli
2	219	100.0	219	US-09-922-480-2	Sequence 2, Appli
3	219	100.0	219	US-09-923-236-2	Sequence 2, Appli
4	9	4.1	2369	US-09-738-626-4434	Sequence 4434, Ap
5	8	3.7	383	US-09-788-345-10	Sequence 10, Appl
6	8	3.7	412	US-09-788-345-12	Sequence 12, Appl
7	8	3.7	428	US-09-906-514-4	Sequence 4, Appl
8	7	3.2	511	US-09-864-761-34590	Sequence 34590, A
9	7	3.2	21	US-10-124-800-15	Sequence 15, Appl
10	7	3.2	27	US-09-974-879-306	Sequence 306, App
11	7	3.2	48	US-09-739-907-177	Sequence 177, App
12	7	3.2	60	US-09-864-761-33554	Sequence 33554, A
13	7	3.2	60	US-09-864-761-33624	Sequence 33624, A
14	7	3.2	60	US-09-864-761-34127	Sequence 34127, A
15	7	3.2	61	US-09-864-761-39187	Sequence 39187, A
16	7	3.2	65	US-09-864-761-36194	Sequence 36194, A
17	7	3.2	88	US-09-764-853-503	Sequence 503, App
18	7	3.2	90	US-09-867-550-738	Sequence 738, App
19	7	3.2	95	US-09-739-907-178	Sequence 178, App

ALIGNMENTS

RESULT 1

US-09-922-469-2

```

; Sequence 2, Application US/09922469
; Patent No. US20020173027A1
; GENERAL INFORMATION:
;
; APPLICANT: Sheppard, Paul O.
; APPLICANT: Adler, David A.
; TITLE OF INVENTION: SECRETED SALIVARY ZSIG63 POLYPEPTIDE
; FILE REFERENCE: 97-71
; CURRENT APPLICATION NUMBER: US/09/922,469
; CURRENT FILING DATE: 2001-08-03
; PRIOR APPLICATION NUMBER: US 60/124,820
; PRIOR FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 219
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-922-469-2

```

Query Match	100.0%;	Score 219;	DB 9;	Length 219;
Best Local Similarity	100.0%;	Pred. No. 2.4e-192;		
Matches 219;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
y	1	MKLLIWACIVCAARKRRPF	IGEDDNDGHPHPSLNIPYGINRLPPPLYRPNVTVP	60
b	1	MKLLIWACIVCAARKRRPF	IGEDDNDGHPHPSLNIPYGINRLPPPLYRPNVTVP	60
y	61	SYPGNTYTDGLPSYPWILLTSPGFPVYHIRGFPLATOLNVPPLPRGRGFVFPVPSRFFSA	120	
b	61	SYPGNTYTDGLPSYPWILLTSPGFPVYHIRGFPLATOLNVPPLPRGRGFVFPVPSRFFSA	120	
y	121	AAAPAAPPTIAEPAAPAAAAPLTATPVAAEPAAAGAPVAAEPAAEPAAEPVAAEPAA	180	
b	121	AAAPAAPPTIAEPAAPAAAAPLTATPVAAEPAAAGAPVAAEPAAEPAAEPVAAEPAA	180	

Sequence 471, Appl
Sequence 720, Appl
Sequence 1234, App
Sequence 476, App
Sequence 2, Appl
Sequence 12052, A
Sequence 4167, App
Sequence 2, Appl
Sequence 2, Appl
Sequence 214, App
Sequence 304, App
Sequence 6, Appl
Sequence 5, Appl
Sequence 20, Appl
Sequence 6, Appl
Sequence 6, Appl
Sequence 10036, App
Sequence 391, App
Sequence 2, Appl
Sequence 391, App
Sequence 21, Appl
Sequence 2, Appl
Sequence 1, Appl
Sequence 272, App
Sequence 373, App
Sequence 5099, App
Sequence 5099, App

```
RESULT 2
US-09-922-480-2
; Sequence 2, Application US/09922480
; Patent No. US20020081701A1
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; APPLICANT: Adler, David A.
; TITLE OF INVENTION: SECRETED SALIVARY ZSIG63 POLYPEPTIDE
; FILE REFERENCE: 97-71
; CURRENT APPLICATION NUMBER: US/09/922,480
; PRIOR FILING DATE: 2001-08-03
; PRIOR APPLICATION NUMBER: US 60/124,820
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 219
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-922-480-2

Query Match      100.0%; Score 219; DB 10; Length 219;
Best Local Similarity 100.0%; Pred. No. 2.4e-192;
Matches 219; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLLWACIVCVAFARRRPFIFIGEDDNDGHLPLHPSLNIPYGINLPPPLYRPNVTVP 60
   |||||||
Db 1 MKLLWACIVCVAFARRRPFIFIGEDDNDGHLPLHPSLNIPYGINLPPPLYRPNVTVP 60

QY 61 SYPGNTYTDGLSPWILTSFGFVYVYHIRGFPLATQLNVPLPGRGFFVPPSRFFSA 120
   |||||||
Db 61 SYPGNTYTDGLSPWILTSFGFVYVYHIRGFPLATQLNVPLPGRGFFVPPSRFFSA 120

QY 121 AAAPAAPPIAAEPAAAPLTATPVAAPAAAGAPVAAEPAAEPVGAEPAAEPVAAEPAA 180
   |||||||
Db 121 AAAPAAPPIAAEPAAAPLTATPVAAPAAAGAPVAAEPAAEPVGAEPAAEPVAAEPAA 180

QY 181 EAPVGVPEAAEPSPAEPATAKPAAPPEPHSPSLEQANQ 219
   |||||||
Db 181 EAPVGVPEAAEPSPAEPATAKPAAPPEPHSPSLEQANQ 219

RESULT 3
US-09-923-236-2
; Sequence 2, Application US/09923236
; Patent No. US20020090677A1
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; APPLICANT: Adler, David A.
; TITLE OF INVENTION: SECRETED SALIVARY ZSIG63 POLYPEPTIDE
; FILE REFERENCE: 97-71
; CURRENT APPLICATION NUMBER: US/09/923,236
; PRIOR FILING DATE: 2001-08-03
; PRIOR APPLICATION NUMBER: US 60/124,820
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 219
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-923-236-2

Query Match      100.0%; Score 219; DB 10; Length 219;
Best Local Similarity 100.0%; Pred. No. 2.4e-192;
Matches 219; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLLWACIVCVAFARRRPFIFIGEDDNDGHLPLHPSLNIPYGINLPPPLYRPNVTVP 60
   |||||||
Db 1 MKLLWACIVCVAFARRRPFIFIGEDDNDGHLPLHPSLNIPYGINLPPPLYRPNVTVP 60

QY 61 SYPGNTYTDGLSPWILTSFGFVYVYHIRGFPLATQLNVPLPGRGFFVPPSRFFSA 120
   |||||||
Db 61 SYPGNTYTDGLSPWILTSFGFVYVYHIRGFPLATQLNVPLPGRGFFVPPSRFFSA 120

QY 121 AAAPAAPPIAAEPAAAPLTATPVAAPAAAGAPVAAEPAAEPVGAEPAAEPVAAEPAA 180
   |||||||
Db 121 AAAPAAPPIAAEPAAAPLTATPVAAPAAAGAPVAAEPAAEPVGAEPAAEPVAAEPAA 180

QY 181 EAPVGVPEAAEPSPAEPATAKPAAPPEPHSPSLEQANQ 219
   |||||||
Db 181 EAPVGVPEAAEPSPAEPATAKPAAPPEPHSPSLEQANQ 219

RESULT 4
US-09-738-626-4434
; Sequence 4434, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 4434
; LENGTH: 2969
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-4434

Query Match      4.1%; Score 9; DB 9; Length 2969;
Best Local Similarity 100.0%; Pred. No. 6.9;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 168 PAAPVAA 176
   |||||||
Db 1663 PAAPVAA 1671

RESULT 5
US-09-788-345-10
; Sequence 10, Application US/09788345
; Patent No. US20020147321A1
; GENERAL INFORMATION:
; APPLICANT: ALONSO BEDATE, Carlos
; APPLICANT: REQUENA ROLANIA, Jose M.
; APPLICANT: SOTO ALVAREZ, Manuel
; TITLE OF INVENTION: CHIMERIC GENE FORMED OF THE DNA SEQUENCES THAT ENCODE THE ANTI
; TITLE OF INVENTION: DETERMINANTS OF FOUR PROTEINS OF L. INFANTUM, USEFUL FOR SERO
; FILE REFERENCE: BEDATE-1A
; CURRENT APPLICATION NUMBER: US/09/788,345
; CURRENT FILING DATE: 2001-02-21
; PRIOR APPLICATION NUMBER: 09/219,306
; PRIOR FILING DATE: 1998-12-23
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 10
; LENGTH: 383
```

; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chimeric
US-09-788-345-10

Query Match 3.7%; Score 8; DB 10; Length 383;
Best Local Similarity 100.0%; Pred. No. 8;
Matches 8; Conservative 0; Mismatches 0; Gaps 0;

QY 120 AAAAPAAP 127
|||||
Db 354 AAAAPAAP 361

RESULT 6

US-09-788-345-12
; Sequence 12, Application US/09788345
; Patent No. US20020147321A1
; GENERAL INFORMATION:
; APPLICANT: ALONSO BEDATE, Carlos
; APPLICANT: REQUENA ROLANIA, Jose M.
; APPLICANT: SOTO ALVAREZ, Manuel
; TITLE OF INVENTION: CHIMERIC GENE FORMED OF THE DNA SEQUENCES THAT ENCODE THE ANTIGEN
; TITLE OF INVENTION: DETERMINANTS OF FOUR PROTEINS OF L. INFANTUM, USEFUL FOR SEROLOG
; TITLE OF INVENTION: CANINE LEISHMANIOSIS AND PROTEIN OBTAINED
; FILE REFERENCE: BEDATE-1A
; CURRENT APPLICATION NUMBER: US/09/788,345
; CURRENT FILING DATE: 2001-02-21
; PRIOR APPLICATION NUMBER: 09/219,306
; PRIOR FILING DATE: 1998-12-23
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 12
; LENGTH: 412
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: recombinant sequence from vectors pQ31 and pMal
US-09-788-345-12

Query Match 3.7%; Score 8; DB 10; Length 412;
Best Local Similarity 100.0%; Pred. No. 8.5;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 120 AAAAPAAP 127
|||||
Db 383 AAAAPAAP 390

RESULT 7

US-09-906-514-4
; Sequence 4, Application US/09906514
; Patent No. US20020170085A1
; GENERAL INFORMATION:
; APPLICANT: Kaeppler, Shawn
; APPLICANT: Springer, Nathan
; APPLICANT: Phillips, Ronald
; TITLE OF INVENTION: Methyl CpG Binding Domain Nucleic Acids from Maize
; FILE REFERENCE: Methybinding
; CURRENT APPLICATION NUMBER: US/09/906,514
; CURRENT FILING DATE: 2001-07-16
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 428
; TYPE: PRT
; ORGANISM: Zea mays
US-09-906-514-4

Query Match 3.7%; Score 8; DB 9; Length 428;
Best Local Similarity 100.0%; Pred. No. 8.9;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 131 AEPAAAP 138
|||||
Db 292 AEPAAAP 299

RESULT 8

US-09-864-761-34590
; Sequence 34590, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Aptomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 34590
; LENGTH: 511
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AB023049.1
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 5.7
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 9.3
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.9
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 7.2
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 5.8
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.8

```
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.1
; OTHER INFORMATION: SWISSPROT HIT: Q02817, EVALUE 3.00e-24
; OTHER INFORMATION: EST_HUMAN HIT: AA487852.1, EVALUE 2.00e-73
US-09-864-761-34590

Query Match          3.7%; Score 8; DB 10; Length 511;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 204 AAPEPHPS 211
Db 95 AAPEPHPS 102
|||||||

RESULT 9
US-10-124-800-15
; Sequence 15, Application US/10124800
; Publication No. US20020194641A1
; GENERAL INFORMATION:
; APPLICANT: Metz, James
; APPLICANT: Barclay, William
; APPLICANT: Flatt, James
; APPLICANT: Kuner, Jerry
; TITLE OF INVENTION: PURA Polyketide Synthase Systems and Uses Thereof
; FILE REFERENCE: 2997-29
; CURRENT APPLICATION NUMBER: US/10/124,800
; CURRENT FILING DATE: 2002-04-16
; PRIOR APPLICATION NUMBER: 09/231,899
; PRIOR FILING DATE: 1999-01-14
; PRIOR APPLICATION NUMBER: 60/284,066
; PRIOR FILING DATE: 2001-04-16
; PRIOR APPLICATION NUMBER: 60/298,796
; PRIOR FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: 60/323,269
; PRIOR FILING DATE: 2001-09-18
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 15
; LENGTH: 21
; TYPE: PRT
; ORGANISM: Schizochytrium sp.
US-10-124-800-15

Query Match          3.2%; Score 7; DB 9; Length 21;
Best Local Similarity 100.0%; Pred. No. 4.2;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 121 AAAPAAP 127
Db 7 AAAPAAP 13
|||||||

RESULT 10
US-09-974-879-306
; Sequence 306, Application US/09974879
; Publication No. US20030028003A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 125 Human Secreted Proteins
; FILE REFERENCE: P2020P2
; CURRENT APPLICATION NUMBER: US/09/974,879
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/239,893
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: US 09/818,683
; PRIOR FILING DATE: 2001-03-28
; PRIOR APPLICATION NUMBER: US 09/305,736
; PRIOR FILING DATE: 1999-05-05
; PRIOR APPLICATION NUMBER: PCT/US98/23435
; PRIOR FILING DATE: 1998-11-04
; PRIOR APPLICATION NUMBER: US 60/064,911
; PRIOR FILING DATE: 1997-11-07
; PRIOR APPLICATION NUMBER: US 60/064,912

Query Match          3.2%; Score 7; DB 10; Length 48;
Best Local Similarity 100.0%; Pred. No. 9.1;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.1
; OTHER INFORMATION: SWISSPROT HIT: Q02817, EVALUE 3.00e-24
; OTHER INFORMATION: EST_HUMAN HIT: AA487852.1, EVALUE 2.00e-73
US-09-864-761-34590

Query Match          3.7%; Score 8; DB 10; Length 511;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 204 AAPEPHPS 211
Db 95 AAPEPHPS 102
|||||||

RESULT 9
US-10-124-800-15
; Sequence 15, Application US/10124800
; Publication No. US20020194641A1
; GENERAL INFORMATION:
; APPLICANT: Metz, James
; APPLICANT: Barclay, William
; APPLICANT: Flatt, James
; APPLICANT: Kuner, Jerry
; TITLE OF INVENTION: PURA Polyketide Synthase Systems and Uses Thereof
; FILE REFERENCE: 2997-29
; CURRENT APPLICATION NUMBER: US/10/124,800
; CURRENT FILING DATE: 2002-04-16
; PRIOR APPLICATION NUMBER: 09/231,899
; PRIOR FILING DATE: 1999-01-14
; PRIOR APPLICATION NUMBER: 60/284,066
; PRIOR FILING DATE: 2001-04-16
; PRIOR APPLICATION NUMBER: 60/298,796
; PRIOR FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: 60/323,269
; PRIOR FILING DATE: 2001-09-18
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 15
; LENGTH: 21
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-974-879-306

Query Match          3.2%; Score 7; DB 9; Length 27;
Best Local Similarity 100.0%; Pred. No. 5.3;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 200 TAKPAAP 206
Db 1 TAKPAAP 7
|||||||

RESULT 11
US-09-739-907-177
; Sequence 177, Application US/09739907
; Patent No. US20010012889A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 36 Human Secreted Proteins
; FILE REFERENCE: P2022P1
; CURRENT APPLICATION NUMBER: US/09/739,907
; CURRENT FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: 09/348,457
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: 60/070,567
; PRIOR FILING DATE: 1998-01-07
; PRIOR APPLICATION NUMBER: 60/070,692
; PRIOR FILING DATE: 1998-01-07
; PRIOR APPLICATION NUMBER: 60/070,704
; PRIOR FILING DATE: 1998-01-07
; PRIOR APPLICATION NUMBER: 60/070,658
; PRIOR FILING DATE: 1998-01-07
; NUMBER OF SEQ ID NOS: 196
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 177
; LENGTH: 48
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-739-907-177

Query Match          3.2%; Score 7; DB 10; Length 48;
Best Local Similarity 100.0%; Pred. No. 9.1;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

QY 102 PPLPPRG 108
| | | | |
Db 42 PPLPPRG 48

RESULT 12

US-09-864-761-33554
; Sequence 33554, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aomico-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 33554
; LENGTH: 60
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2.1
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 2.3
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.96
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.6

; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.4
; OTHER INFORMATION: EST_HUMAN HIT: L25488.1, EVALUE 3.00e+00
; OTHER INFORMATION: SWISSPROT HIT: O95104, EVALUE 9.00e-07
US-09-864-761-33554

Query Match 3.2%; Score 7; DB 10; Length 60;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 120 AAAAPAA 126
| | | | |
Db 45 AAAAPAA 51

RESULT 13

US-09-864-761-33624
; Sequence 33624, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
; FILE REFERENCE: Aomico-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 33624
; LENGTH: 60
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AP000213.1
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 3.2

```
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.9
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.6
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2.2
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 4.1
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.9
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.7
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.1
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.9
; OTHER INFORMATION: EST_HUMAN HIT: L25488.1, EVALUE 3.00e+00
; OTHER INFORMATION: SWISSPROT HIT: Q95104, EVALUE 9.00e-07
US-09-864-761-33624
```

Query Match 3.2% Score 7; DB 10; Length 60;

Best Local Similarity 100.0%; Pred. No. 11;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 120 AAAAPAA 126

Db 45 AAAAPAA 51

RESULT 14

```
US-09-864-761-34127
; Sequence 34127, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aemica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
```

```
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 34127
; LENGTH: 60
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AP000135.1
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 3.3
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3.6
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3.7
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 4.1
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 3.1
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 3.4
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.5
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3.6
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.4
; OTHER INFORMATION: EST_HUMAN HIT: L25488.1, EVALUE 3.00e+00
; OTHER INFORMATION: SWISSPROT HIT: Q95104, EVALUE 9.00e-07
US-09-864-761-34127
```

Query Match 3.2% Score 7; DB 10; Length 60;

Best Local Similarity 100.0%; Pred. No. 11;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 120 AAAAPAA 126

Db 45 AAAAPAA 51

RESULT 15

```
US-09-864-761-39187
; Sequence 39187, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
; FILE REFERENCE: Aemica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
```

; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
; SEQ ID NO 39187
; LENGTH: 61
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC005015.2
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 4.1
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 3.7
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 3.7
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.6
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 4.3
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 5.1
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 4.3
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3.5
; OTHER INFORMATION: EST_HUMAN HIT: AW503503.1, EVALUE 3.00e-06
US-09-864-761-39187

Query Match 3.28; Score 7; DB 10; Length 61;
Best Local Similarity 100.0%; Pred.No.11;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 119 SAAAPAA 125
Db 30 SAAAPAA 36

Search completed: February 12, 2003, 14:03:47
Job time : 16 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2003, 13:57:55 ; Search time 141 Seconds
(without alignments)
1001.394 Million cell updates

Title: US-09-923-236-2

perfect score: 219
Sequence: 1 MKLLWACIVCAFAKRRF.....TAKPAAPHPSPSLEOANO 219

Scoring table: OLIGO

Searched: 4569144 seqs, 644733110 residues

Total number of hits satisfying chosen parameters: 4569144

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : Pending_Patents_AA_Main:*

1:	/cgn2_6/ptodata/1/paa/pctus_COMB.pcp:*
2:	/cgn2_6/ptodata/1/paa/US06_COMB.pcp:*
3:	/cgn2_6/ptodata/1/paa/US07_COMB.pcp:*
4:	/cgn2_6/ptodata/1/paa/US080_COMB.pcp:*
5:	/cgn2_6/ptodata/1/paa/US081_COMB.pcp:*
6:	/cgn2_6/ptodata/1/paa/US082_COMB.pcp:*
7:	/cgn2_6/ptodata/1/paa/US083_COMB.pcp:*
8:	/cgn2_6/ptodata/1/paa/US084_COMB.pcp:*
9:	/cgn2_6/ptodata/1/paa/US085_COMB.pcp:*
10:	/cgn2_6/ptodata/1/paa/US086_COMB.pcp:*
11:	/cgn2_6/ptodata/1/paa/US087_COMB.pcp:*
12:	/cgn2_6/ptodata/1/paa/US088_COMB.pcp:*
13:	/cgn2_6/ptodata/1/paa/US089_COMB.pcp:*
14:	/cgn2_6/ptodata/1/paa/US090_COMB.pcp:*
15:	/cgn2_6/ptodata/1/paa/US091_COMB.pcp:*
16:	/cgn2_6/ptodata/1/paa/US092_COMB.pcp:*
17:	/cgn2_6/ptodata/1/paa/US093_COMB.pcp:*
18:	/cgn2_6/ptodata/1/paa/US094_COMB.pcp:*
19:	/cgn2_6/ptodata/1/paa/US095_COMB.pcp:*
20:	/cgn2_6/ptodata/1/paa/US096_COMB.pcp:*
21:	/cgn2_6/ptodata/1/paa/US097_COMB.pcp:*
22:	/cgn2_6/ptodata/1/paa/US098_COMB.pcp:*
23:	/cgn2_6/ptodata/1/paa/US099_COMB.pcp:*
24:	/cgn2_6/ptodata/1/paa/US100_COMB.pcp:*
25:	/cgn2_6/ptodata/1/paa/US101_COMB.pcp:*
26:	/cgn2_6/ptodata/1/paa/US102_COMB.pcp:*
27:	/cgn2_6/ptodata/1/paa/US103_COMB.pcp:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query q			DB	ID	Description
	Score	Match	Length			
1	219	100.0	219	1	PCT-US02-08123-1081	Sequence 1081, Ap
2	219	100.0	219	1	PCT-US02-08277-723	Sequence 723, App
3	219	100.0	219	1	PCT-US02-08278-1040	Sequence 1040, Ap
4	219	100.0	219	23	US-09-922-469-2	Sequence 2, Appli
5	219	100.0	219	23	US-09-922-480-2	Sequence 2, Appli
6	219	100.0	219	23	US-09-923-236-2	Sequence 2, Appli

7	219	100.0	219	24	US-10-009-328-29	Sequence 29, Appl
8	219	100.0	219	27	US-60-162-287-4	Sequence 4, Appl
9	150	68.5	220	1	PCT-US02-08123-1632	Sequence 1632, Ap
10	150	68.5	220	1	PCT-US03-08277-1086	Sequence 1086, Ap
11	150	68.5	220	1	PCT-US03-08278-1563	Sequence 1563, Ap
12	150	68.5	221	1	PCT-US98-22376-190	Sequence 190, App
13	150	68.5	221	16	US-09-296-622-192	Sequence 192, App
14	150	68.5	221	26	US-10-231-417-192	Sequence 192, App
15	140	63.9	219	1	PCT-US98-13813-13	Sequence 10, Appl
16	140	63.9	219	13	US-08-958-304-10	Sequence 10, Appl
17	140	63.9	219	21	US-09-746-783-106	Sequence 106, App
18	126	57.5	143	22	US-09-834-366-18049	Sequence 18049, A
19	126	57.5	143	27	US-60-197-873-18049	Sequence 18049, A
20	125	57.1	143	20	US-09-621-976-5226	Sequence 5226, Ap
21	125	57.1	143	27	US-60-147-499-5226	Sequence 5226, Ap
22	112	51.1	112	12	US-08-887-195-14	Sequence 14, Appl
23	109	49.8	120	20	US-09-621-976-3905	Sequence 3905, Ap
24	109	49.8	120	27	US-60-147-499-3905	Sequence 3905, Ap
25	58	26.5	91	22	US-09-834-366-13475	Sequence 13475, A
26	58	26.5	91	27	US-60-197-873-13475	Sequence 13475, A
27	10	4.6	111	21	US-09-708-427-72642	Sequence 72642, A
28	9	4.1	75	20	US-09-614-150-19398	Sequence 19398, A
29	9	4.1	75	27	US-60-167-217-10318	Sequence 10318, A
30	9	4.1	75	27	US-60-167-217-19534	Sequence 19534, A
31	9	4.1	75	27	US-60-173-464-19568	Sequence 19568, A
32	9	4.1	75	27	US-60-191-637-19457	Sequence 19457, A
33	9	4.1	75	27	US-60-191-681-15371	Sequence 15371, A
34	9	4.1	334	16	US-09-270-767-41733	Sequence 41733, A
35	9	4.1	697	27	US-60-161-932-24082	Sequence 24082, Ap
36	9	4.1	728	20	US-09-614-150-23082	Sequence 23082, A
37	9	4.1	728	21	US-09-791-537-144373	Sequence 144373, A
38	9	4.1	728	27	US-60-173-464-19004	Sequence 19004, A
39	9	4.1	728	27	US-60-191-637-23165	Sequence 23165, A
40	9	4.1	728	27	US-60-191-681-18263	Sequence 18263, A
41	9	4.1	864	21	US-09-791-537-48360	Sequence 48360, A
42	9	4.1	1874	20	US-09-602-787A-46	Sequence 46, Appl
43	9	4.1	2369	21	US-09-738-626-4434	Sequence 4434, Ap
44	8	3.7	65	1	PCT-US99-22855-2421	Sequence 2421, Ap
45	8	3.7	65	19	US-09-513-996A-32499	Sequence 32499, A

ALIGNMENTS

RESULT 1

```

PCT-US02-08123-1081
; Sequence 1081, Application PC/TUS0208123
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc.
; TITLE OF INVENTION: Human Secreted Proteins
; FILE REFERENCE: PS904PCT
; CURRENT APPLICATION NUMBER: PCT/US02/08123
; CURRENT FILING DATE: 2002-03-19
; PRIOR APPLICATION NUMBER: US 60/331,287
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/306,171
; PRIOR FILING DATE: 2001-07-19
; PRIOR APPLICATION NUMBER: US 60/277,340
; PRIOR FILING DATE: 2001-03-21
; NUMBER OF SEQ ID NOS: 2048
; SOFTWARE: Patentn Ver. 2.0
; SEQ ID NO 1081
; LENGTH: 219
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US02-08123-1081

```

Query Match 100.0%; Score 219; DB 1; Length 219;

Best Local Similarity 100.00%, Fied. NO. 0E-133,
Matches 219: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLLWACIVCVAFARKKRRFPFIGEDDNDGDHPLPSLNIPYGINLPPLYYRPPVNTVP 60
QY |||||

```
Db 1 MKLLWACIVCAFAKRRFFIGEDDNDGHLHPSLNIPYGINLPPPLLYRPVNTVP 60
Qy 61 SYPGNTYTDGLSPYVWILTSFGFPYVYHIRGFPLATQLNVPLPPRGFPFVPPSRFFSA 120
Db 61 SYPGNTYTDGLSPYVWILTSFGFPYVYHIRGFPLATQLNVPLPPRGFPFVPPSRFFSA 120
Qy 121 AAAPAAPPIAAEPAAPAAAPLTATPVAAEPAAGAPVAAEPAAPVAAEPAAPVAAEPAAP 180
Db 121 AAAPAAPPIAAEPAAPAAAPLTATPVAAEPAAGAPVAAEPAAPVAAEPAAPVAAEPAAP 180
Qy 181 EAPVGEPAEPAEPPSPAEPAATAKPAAPPEHPSPSLEQANQ 219
Db 181 EAPVGEPAEPAEPPSPAEPAATAKPAAPPEHPSPSLEQANQ 219

RESULT 2
PCT-US02-08277-723
; Sequence 723, Application PC/TUS0208277
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc.
; TITLE OF INVENTION: Human Secreted Proteins
; FILE REFERENCE: PS907PCT
; CURRENT APPLICATION NUMBER: PCT/US02/08277
; CURRENT FILING DATE: 2002-03-19
; PRIOR APPLICATION NUMBER: US 60/331,287
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/306,171
; PRIOR FILING DATE: 2001-07-19
; PRIOR APPLICATION NUMBER: US 60/277,340
; PRIOR FILING DATE: 2001-03-21
; NUMBER OF SEQ ID NOS: 1357
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 723
; LENGTH: 219
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US02-08277-723

Query Match 100.0%; Score 219; DB 1; Length 219;
Best Local Similarity 100.0%; Pred. No. 6e-193;
Matches 219; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MKLLWACIVCAFAKRRFFIGEDDNDGHLHPSLNIPYGINLPPPLLYRPVNTVP 60
Db 1 MKLLWACIVCAFAKRRFFIGEDDNDGHLHPSLNIPYGINLPPPLLYRPVNTVP 60
Qy 61 SYPGNTYTDGLSPYVWILTSFGFPYVYHIRGFPLATQLNVPLPPRGFPFVPPSRFFSA 120
Db 61 SYPGNTYTDGLSPYVWILTSFGFPYVYHIRGFPLATQLNVPLPPRGFPFVPPSRFFSA 120
Qy 121 AAAPAAPPIAAEPAAPAAAPLTATPVAAEPAAGAPVAAEPAAPVAAEPAAPVAAEPAAP 180
Db 121 AAAPAAPPIAAEPAAPAAAPLTATPVAAEPAAGAPVAAEPAAPVAAEPAAPVAAEPAAP 180
Qy 181 EAPVGEPAEPAEPPSPAEPAATAKPAAPPEHPSPSLEQANQ 219
Db 181 EAPVGEPAEPAEPPSPAEPAATAKPAAPPEHPSPSLEQANQ 219

RESULT 3
PCT-US02-08278-1040
; Sequence 1040, Application PC/TUS0208278
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc.
; TITLE OF INVENTION: Human Secreted Proteins
; FILE REFERENCE: PS902PCT
; CURRENT APPLICATION NUMBER: PCT/US02/08278
; CURRENT FILING DATE: 2002-03-19
; PRIOR APPLICATION NUMBER: US 60/331,287
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/306,171
; PRIOR FILING DATE: 2001-07-19
; PRIOR APPLICATION NUMBER: US 60/277,340

Query Match 100.0%; Score 219; DB 1; Length 219;
Best Local Similarity 100.0%; Pred. No. 6e-193;
Matches 219; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MKLLWACIVCAFAKRRFFIGEDDNDGHLHPSLNIPYGINLPPPLLYRPVNTVP 60
Db 1 MKLLWACIVCAFAKRRFFIGEDDNDGHLHPSLNIPYGINLPPPLLYRPVNTVP 60
Qy 61 SYPGNTYTDGLSPYVWILTSFGFPYVYHIRGFPLATQLNVPLPPRGFPFVPPSRFFSA 120
Db 61 SYPGNTYTDGLSPYVWILTSFGFPYVYHIRGFPLATQLNVPLPPRGFPFVPPSRFFSA 120
Qy 121 AAAPAAPPIAAEPAAPAAAPLTATPVAAEPAAGAPVAAEPAAPVAAEPAAPVAAEPAAP 180
Db 121 AAAPAAPPIAAEPAAPAAAPLTATPVAAEPAAGAPVAAEPAAPVAAEPAAPVAAEPAAP 180
Qy 181 EAPVGEPAEPAEPPSPAEPAATAKPAAPPEHPSPSLEQANQ 219
Db 181 EAPVGEPAEPAEPPSPAEPAATAKPAAPPEHPSPSLEQANQ 219

RESULT 4
US-09-922-469-2
; Sequence 2, Application US/09922469
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul A.
; TITLE OF INVENTION: SECRETED SALIVARY ZS1G63 POLYPEPTIDE
; FILE REFERENCE: 97-71
; CURRENT APPLICATION NUMBER: US/09/922,469
; CURRENT FILING DATE: 2001-08-03
; PRIOR APPLICATION NUMBER: US 60/124,820
; PRIOR FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 219
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-922-469-2

Query Match 100.0%; Score 219; DB 23; Length 219;
Best Local Similarity 100.0%; Pred. No. 6e-193;
Matches 219; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MKLLWACIVCAFAKRRFFIGEDDNDGHLHPSLNIPYGINLPPPLLYRPVNTVP 60
Db 1 MKLLWACIVCAFAKRRFFIGEDDNDGHLHPSLNIPYGINLPPPLLYRPVNTVP 60
Qy 61 SYPGNTYTDGLSPYVWILTSFGFPYVYHIRGFPLATQLNVPLPPRGFPFVPPSRFFSA 120
Db 61 SYPGNTYTDGLSPYVWILTSFGFPYVYHIRGFPLATQLNVPLPPRGFPFVPPSRFFSA 120
Qy 121 AAAPAAPPIAAEPAAPAAAPLTATPVAAEPAAGAPVAAEPAAPVAAEPAAPVAAEPAAP 180
Db 121 AAAPAAPPIAAEPAAPAAAPLTATPVAAEPAAGAPVAAEPAAPVAAEPAAPVAAEPAAP 180
Qy 181 EAPVGEPAEPAEPPSPAEPAATAKPAAPPEHPSPSLEQANQ 219
Db 181 EAPVGEPAEPAEPPSPAEPAATAKPAAPPEHPSPSLEQANQ 219

RESULT 5
US-09-922-480-2
```

```
; PRIOR FILING DATE: 2001-03-21
; NUMBER OF SEQ ID NOS: 1988
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1040
; LENGTH: 219
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US02-08278-1040

Query Match 100.0%; Score 219; DB 1; Length 219;
Best Local Similarity 100.0%; Pred. No. 6e-193;
Matches 219; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MKLLWACIVCAFAKRRFFIGEDDNDGHLHPSLNIPYGINLPPPLLYRPVNTVP 60
Db 1 MKLLWACIVCAFAKRRFFIGEDDNDGHLHPSLNIPYGINLPPPLLYRPVNTVP 60
Qy 61 SYPGNTYTDGLSPYVWILTSFGFPYVYHIRGFPLATQLNVPLPPRGFPFVPPSRFFSA 120
Db 61 SYPGNTYTDGLSPYVWILTSFGFPYVYHIRGFPLATQLNVPLPPRGFPFVPPSRFFSA 120
Qy 121 AAAPAAPPIAAEPAAPAAAPLTATPVAAEPAAGAPVAAEPAAPVAAEPAAPVAAEPAAP 180
Db 121 AAAPAAPPIAAEPAAPAAAPLTATPVAAEPAAGAPVAAEPAAPVAAEPAAPVAAEPAAP 180
Qy 181 EAPVGEPAEPAEPPSPAEPAATAKPAAPPEHPSPSLEQANQ 219
Db 181 EAPVGEPAEPAEPPSPAEPAATAKPAAPPEHPSPSLEQANQ 219

RESULT 4
US-09-922-469-2
; Sequence 2, Application US/09922469
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul A.
; TITLE OF INVENTION: SECRETED SALIVARY ZS1G63 POLYPEPTIDE
; FILE REFERENCE: 97-71
; CURRENT APPLICATION NUMBER: US/09/922,469
; CURRENT FILING DATE: 2001-08-03
; PRIOR APPLICATION NUMBER: US 60/124,820
; PRIOR FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 219
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-922-469-2

Query Match 100.0%; Score 219; DB 23; Length 219;
Best Local Similarity 100.0%; Pred. No. 6e-193;
Matches 219; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MKLLWACIVCAFAKRRFFIGEDDNDGHLHPSLNIPYGINLPPPLLYRPVNTVP 60
Db 1 MKLLWACIVCAFAKRRFFIGEDDNDGHLHPSLNIPYGINLPPPLLYRPVNTVP 60
Qy 61 SYPGNTYTDGLSPYVWILTSFGFPYVYHIRGFPLATQLNVPLPPRGFPFVPPSRFFSA 120
Db 61 SYPGNTYTDGLSPYVWILTSFGFPYVYHIRGFPLATQLNVPLPPRGFPFVPPSRFFSA 120
Qy 121 AAAPAAPPIAAEPAAPAAAPLTATPVAAEPAAGAPVAAEPAAPVAAEPAAPVAAEPAAP 180
Db 121 AAAPAAPPIAAEPAAPAAAPLTATPVAAEPAAGAPVAAEPAAPVAAEPAAPVAAEPAAP 180
Qy 181 EAPVGEPAEPAEPPSPAEPAATAKPAAPPEHPSPSLEQANQ 219
Db 181 EAPVGEPAEPAEPPSPAEPAATAKPAAPPEHPSPSLEQANQ 219

RESULT 5
US-09-922-480-2
```



```
; APPLICANT: Lu, Dyung Aina M.
; TITLE OF INVENTION: TRANSPORT PROTEINS
; FILE REFERENCE: PF-0748 P
; CURRENT APPLICATION NUMBER: US/60/162,287
; CURRENT FILING DATE: 1999-10-28
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: PERL Program
; SEQ ID NO 4
; LENGTH: 219
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; OTHER INFORMATION: Incyte ID No: I325518CD1
US-60-162-287-4

Query Match          100.0%; Score 219; DB 27; Length 219;
Best Local Similarity 100.0%; Pred. No. 6e-193;
Matches 219; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLLWACIVCVAFARKRRPFPIGEDDDGHPHPSLNIPYGINLPPPLYRPNVTVP 60
   ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1 MKLLWACIVCVAFARKRRPFPIGEDDDGHPHPSLNIPYGINLPPPLYRPNVTVP 60
   ||||||||||||||||||||||||||||||||||||||||||||||||||||||

QY 61 SYPGNTYTDGLSPWILTSFGFPYVYHIRGFPLATQLNVPLPRGPFVPPSRFFSA 120
   ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 61 SYPGNTYTDGLSPWILTSFGFPYVYHIRGFPLATQLNVPLPRGPFVPPSRFFSA 120
   ||||||||||||||||||||||||||||||||||||||||||||||||||||||

QY 121 AAAPAAPPIAAEPAAAAPLTATPVAAEPAA 180
   ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 121 AAAPAAPPIAAEPAAAAPLTATPVAAEPAA 180
   ||||||||||||||||||||||||||||||||||||||||||||||||||||||

QY 181 EAPVGVPEAAEESPAAEPATAPAAPEHPSPSLEQANQ 219
   ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 181 EAPVGVPEAAEESPAAEPATAPAAPEHPSPSLEQANQ 219
   ||||||||||||||||||||||||||||||||||||||||||||||||||||||

RESULT 9
PCT-US02-08123-1632
; Sequence 1632, Application PC/TUS0208123
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc.
; TITLE OF INVENTION: Human Secreted Proteins
; FILE REFERENCE: PS904PCT
; CURRENT APPLICATION NUMBER: PCT/US02/08123
; CURRENT FILING DATE: 2002-03-19
; PRIOR APPLICATION NUMBER: US 60/331,287
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/306,171
; PRIOR FILING DATE: 2001-07-19
; PRIOR APPLICATION NUMBER: US 60/277,340
; NUMBER OF SEQ ID NOS: 2048
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1632
; LENGTH: 220
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (159)
; OTHER INFORMATION: Xaa equals any amino acid
PCT-US02-08123-1632

Query Match          68.5%; Score 150; DB 1; Length 220;
Best Local Similarity 100.0%; Pred. No. 2.3e-129;
Matches 150; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLLWACIVCVAFARKRRPFPIGEDDDGHPHPSLNIPYGINLPPPLYRPNVTVP 60
   ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1 MKLLWACIVCVAFARKRRPFPIGEDDDGHPHPSLNIPYGINLPPPLYRPNVTVP 60
   ||||||||||||||||||||||||||||||||||||||||||||||||||||||

QY 61 SYPGNTYTDGLSPWILTSFGFPYVYHIRGFPLATQLNVPLPRGPFVPPSRFFSA 120
   ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 61 SYPGNTYTDGLSPWILTSFGFPYVYHIRGFPLATQLNVPLPRGPFVPPSRFFSA 120
   ||||||||||||||||||||||||||||||||||||||||||||||||||||||

QY 121 AAAPAAPPIAAEPAAAAPLTATPVAAEPAA 150
   ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 121 AAAPAAPPIAAEPAAAAPLTATPVAAEPAA 150
   ||||||||||||||||||||||||||||||||||||||||||||||||||||||

RESULT 11
PCT-US02-08278-1563
; Sequence 1563, Application PC/TUS0208278
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc.
; TITLE OF INVENTION: Human Secreted Proteins
; FILE REFERENCE: PS902PCT
; CURRENT APPLICATION NUMBER: PCT/US02/08278
; CURRENT FILING DATE: 2002-03-19
; PRIOR APPLICATION NUMBER: US 60/331,287
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/306,171
; PRIOR FILING DATE: 2001-07-19
; PRIOR APPLICATION NUMBER: US 60/277,340
; NUMBER OF SEQ ID NOS: 1988
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1563
; LENGTH: 220
; TYPE: PRT
; ORGANISM: Homo sapiens
```

```
|||||
Db 61 SYPGNTYTDGLSPWILTSFGFPYVYHIRGFPLATQLNVPLPRGPFVPPSRFFSA 120
   ||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 121 AAAPAAPPIAAEPAAAAPLTATPVAAEPAA 150
   ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 121 AAAPAAPPIAAEPAAAAPLTATPVAAEPAA 150
   ||||||||||||||||||||||||||||||||||||||||||||||||||||||

RESULT 10
PCT-US02-08277-1086
; Sequence 1086, Application PC/TUS0208277
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc.
; TITLE OF INVENTION: Human Secreted Proteins
; FILE REFERENCE: PS907PCT
; CURRENT APPLICATION NUMBER: PCT/US02/08277
; CURRENT FILING DATE: 2002-03-19
; PRIOR APPLICATION NUMBER: US 60/331,287
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/306,171
; PRIOR FILING DATE: 2001-07-19
; PRIOR APPLICATION NUMBER: US 60/277,340
; PRIOR FILING DATE: 2001-03-21
; NUMBER OF SEQ ID NOS: 1357
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1086
; LENGTH: 220
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (159)
; OTHER INFORMATION: Xaa equals any amino acid
PCT-US02-08277-1086

Query Match          68.5%; Score 150; DB 1; Length 220;
Best Local Similarity 100.0%; Pred. No. 2.3e-129;
Matches 150; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLLWACIVCVAFARKRRPFPIGEDDDGHPHPSLNIPYGINLPPPLYRPNVTVP 60
   ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1 MKLLWACIVCVAFARKRRPFPIGEDDDGHPHPSLNIPYGINLPPPLYRPNVTVP 60
   ||||||||||||||||||||||||||||||||||||||||||||||||||||||

QY 61 SYPGNTYTDGLSPWILTSFGFPYVYHIRGFPLATQLNVPLPRGPFVPPSRFFSA 120
   ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 61 SYPGNTYTDGLSPWILTSFGFPYVYHIRGFPLATQLNVPLPRGPFVPPSRFFSA 120
   ||||||||||||||||||||||||||||||||||||||||||||||||||||||

QY 121 AAAPAAPPIAAEPAAAAPLTATPVAAEPAA 150
   ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 121 AAAPAAPPIAAEPAAAAPLTATPVAAEPAA 150
   ||||||||||||||||||||||||||||||||||||||||||||||||||||||

RESULT 11
PCT-US02-08278-1563
; Sequence 1563, Application PC/TUS0208278
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc.
; TITLE OF INVENTION: Human Secreted Proteins
; FILE REFERENCE: PS902PCT
; CURRENT APPLICATION NUMBER: PCT/US02/08278
; CURRENT FILING DATE: 2002-03-19
; PRIOR APPLICATION NUMBER: US 60/331,287
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/306,171
; PRIOR FILING DATE: 2001-07-19
; PRIOR APPLICATION NUMBER: US 60/277,340
; NUMBER OF SEQ ID NOS: 1988
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1563
; LENGTH: 220
; TYPE: PRT
; ORGANISM: Homo sapiens
```

```
;
; NAME/KEY: SITE
; LOCATION: (159)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (221)
; OTHER INFORMATION: Xaa equals stop translation
PCT-US98-22376-190

Query Match      68.5%  Score 150; DB 1; Length 221;
Best Local Similarity 100.0%; Pred. No. 2.3e-129;
Matches 150; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLLWACIVCAVAFKRRFFIGEDDNDGHPHPSLNIPYGINLPPPLYRPNVTVP 60
Db 1 MKLLWACIVCAVAFKRRFFIGEDDNDGHPHPSLNIPYGINLPPPLYRPNVTVP 60
QY 61 SYPGNTYTDGLPSYPWLTSPGFPYVYHGRGFPPLATQLNVPLPPRGFPVPPSRFFSA 120
Db 61 SYPGNTYTDGLPSYPWLTSPGFPYVYHGRGFPPLATQLNVPLPPRGFPVPPSRFFSA 120
QY 121 AAAPAAPPIAAEPAAAPLTATPVAAEPAA 150
Db 121 AAAPAAPPIAAEPAAAPLTATPVAAEPAA 150

RESULT 13
US-09-296-622-192
; Sequence 192, Application US/09296622
; GENERAL INFORMATION:
; APPLICANT: Feng et al.
; TITLE OF INVENTION: 148 Human Secreted Proteins
; FILE REFERENCE: P2019P1
; CURRENT APPLICATION NUMBER: US/09/296,622
; EARLIER FILING DATE: 1999-04-23
; EARLIER APPLICATION NUMBER: PCT/US98/22376
; EARLIER FILING DATE: 1998-10-23
; EARLIER APPLICATION NUMBER: 60/063,099
; EARLIER FILING DATE: 1997-10-24
; EARLIER APPLICATION NUMBER: 60/063,088
; EARLIER FILING DATE: 1997-10-24
; EARLIER APPLICATION NUMBER: 60/063,100
; EARLIER FILING DATE: 1997-10-24
; EARLIER APPLICATION NUMBER: 60/063,387
; EARLIER FILING DATE: 1997-10-24
; EARLIER APPLICATION NUMBER: 60/063,148
; EARLIER FILING DATE: 1997-10-24
; EARLIER APPLICATION NUMBER: 60/063,386
; EARLIER FILING DATE: 1997-10-24
; EARLIER APPLICATION NUMBER: 60/062,784
; EARLIER FILING DATE: 1997-10-24
; EARLIER APPLICATION NUMBER: 60/063,091
; EARLIER FILING DATE: 1997-10-24
; EARLIER APPLICATION NUMBER: 60/063,090
; EARLIER FILING DATE: 1997-10-24
; EARLIER APPLICATION NUMBER: 60/063,089
; EARLIER FILING DATE: 1997-10-24
; EARLIER APPLICATION NUMBER: 60/063,092
; EARLIER FILING DATE: 1997-10-24
; EARLIER APPLICATION NUMBER: 60/063,111
; EARLIER FILING DATE: 1997-10-24
; EARLIER APPLICATION NUMBER: 60/063,101
; EARLIER FILING DATE: 1997-10-24
; EARLIER APPLICATION NUMBER: 60/063,109
; EARLIER FILING DATE: 1997-10-24
; EARLIER APPLICATION NUMBER: 60/063,110
; EARLIER FILING DATE: 1997-10-24
; EARLIER APPLICATION NUMBER: 60/063,098
; EARLIER FILING DATE: 1997-10-24
; EARLIER APPLICATION NUMBER: 60/063,097
; EARLIER FILING DATE: 1997-10-24
; EARLIER APPLICATION NUMBER: 60/063,100
; EARLIER FILING DATE: 1997-10-24
; EARLIER APPLICATION NUMBER: 60/063,098
; EARLIER FILING DATE: 1997-10-24
; NUMBER OF SEQ ID NOS: 492
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 190
; LENGTH: 221
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
```

```
; SEQ ID NO 192
; LENGTH: 221
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (159)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (221)
; OTHER INFORMATION: Xaa equals stop translation
US-09-296-622-192

Query Match          68.5%; Score 150; DB 16; Length 221;
Best Local Similarity 100.0%; Pred. No. 2.3e-129;
Matches 150; Conserved 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLLWACIVCVAFARRRPFIFGDDNDGHLHPSLNIPYGINLPPPLYRPNVTVP 60
Db 1 MKLLWACIVCVAFARRRPFIFGDDNDGHLHPSLNIPYGINLPPPLYRPNVTVP 60

QY 61 SYPGNTYTDGLPSYPWILTSFGFPVYHIRGFPPLATQLNVPLPRGFPFVPPSRFFSA 120
Db 61 SYPGNTYTDGLPSYPWILTSFGFPVYHIRGFPPLATQLNVPLPRGFPFVPPSRFFSA 120

QY 121 AAAPAAPPIAAEPAAAPLTATPVAAEPAA 150
Db 121 AAAPAAPPIAAEPAAAPLTATPVAAEPAA 150

RESULT 14
US-10-231-417-192
; Sequence 192, Application US/10231417
; GENERAL INFORMATION:
; APPLICANT: Feng et al.
; TITLE OF INVENTION: 148 Human Secreted Proteins
; FILE REFERENCE: P2019P1
; CURRENT APPLICATION NUMBER: US/10/231,417
; PRIOR FILING DATE: 2002-08-30
; PRIOR APPLICATION NUMBER: US/09/296,622
; FILING DATE: 1999-04-23
; NUMBER OF SEQ ID NOS: 619
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 192
; LENGTH: 221
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (159)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (221)
; OTHER INFORMATION: Xaa equals stop translation
US-10-231-417-192

Query Match          68.5%; Score 150; DB 26; Length 221;
Best Local Similarity 100.0%; Pred. No. 2.3e-129;
Matches 150; Conserved 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLLWACIVCVAFARRRPFIFGDDNDGHLHPSLNIPYGINLPPPLYRPNVTVP 60
Db 1 MKLLWACIVCVAFARRRPFIFGDDNDGHLHPSLNIPYGINLPPPLYRPNVTVP 60

QY 61 SYPGNTYTDGLPSYPWILTSFGFPVYHIRGFPPLATQLNVPLPRGFPFVPPSRFFSA 120
Db 61 SYPGNTYTDGLPSYPWILTSFGFPVYHIRGFPPLATQLNVPLPRGFPFVPPSRFFSA 120

QY 121 AAAPAAPPIAAEPAAAPLTATPVAAEPAA 150
Db 121 AAAPAAPPIAAEPAAAPLTATPVAAEPAA 150
```

```
RESULT 15
PCT-US98-13813-10
; Sequence 10, Application PC/TUS9813813
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John M.
; APPLICANT: LaValle, Edward R.
; APPLICANT: Racie, Lisa A.
; APPLICANT: Treacy, Maurice
; APPLICANT: Spaulding, Vikki
; APPLICANT: Agostino, Michael J.
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
; TITLE OF INVENTION: ENCODING THEM
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: MA
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US98/13813
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Sprunger, Suzanne A.
; REGISTRATION NUMBER: 41,323
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 498-8284
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 219 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US98-13813-10

Query Match          63.9%; Score 140; DB 1; Length 219;
Best Local Similarity 100.0%; Pred. No. 3.7e-120;
Matches 140; Conserved 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLLWACIVCVAFARRRPFIFGDDNDGHLHPSLNIPYGINLPPPLYRPNVTVP 60
Db 1 MKLLWACIVCVAFARRRPFIFGDDNDGHLHPSLNIPYGINLPPPLYRPNVTVP 60

QY 61 SYPGNTYTDGLPSYPWILTSFGFPVYHIRGFPPLATQLNVPLPRGFPFVPPSRFFSA 120
Db 61 SYPGNTYTDGLPSYPWILTSFGFPVYHIRGFPPLATQLNVPLPRGFPFVPPSRFFSA 120

QY 121 AAAPAAPPIAAEPAAAPLT 140
Db 121 AAAPAAPPIAAEPAAAPLT 140

Search completed: February 12, 2003, 14:02:43
Job time : 142 secs
```

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2003, 13:58:25 ; Search time 24 Seconds
(without alignments)
764.561 Million cell updates/sec

Title: US-09-923-236-2

Perfect score: 219

Sequence: 1 MKLLMVICVFAFKRRF.....TAKPAPEPHSPSLEQANQ 219

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 425801 seqs, 83787487 residues

Word size : 0

Total number of hits satisfying chosen parameters: 425801

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Pending_Patents_AA_New.*

- 1: /cgn2_6/ptodata/1/paa/PCT_NEW_COMB.pep.*
- 2: /cgn2_6/ptodata/1/paa/US06_NEW_COMB.pep.*
- 3: /cgn2_6/ptodata/1/paa/US07_NEW_COMB.pep.*
- 4: /cgn2_6/ptodata/1/paa/US08_NEW_COMB.pep.*
- 5: /cgn2_6/ptodata/1/paa/US09_NEW_COMB.pep.*
- 6: /cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep.*
- 7: /cgn2_6/ptodata/1/paa/US60_NEW_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	8	3.7	121	6 US-10-218-140-1314	Sequence 1314, Ap
2	8	3.7	383	6 US-10-337-312-10	Sequence 10, Appl
3	8	3.7	412	6 US-10-337-312-12	Sequence 12, Appl
4	8	3.7	428	6 US-10-198-723-14	Sequence 14, Appl
5	8	3.7	511	6 US-10-203-138A-11635	Sequence 11635, A
6	8	3.7	512	6 US-10-263-929-103	Sequence 103, App
7	7	3.2	51	5 US-09-724-676-91213	Sequence 91213, A
8	7	3.2	51	5 US-09-724-676A-91213	Sequence 91213, A
9	7	3.2	53	1 PCT-US02-32727-19524	Sequence 19524, A
10	7	3.2	53	6 US-10-057-498-19524	Sequence 19524, A
11	7	3.2	59	1 PCT-US02-32727-12349	Sequence 12349, A
12	7	3.2	59	6 US-10-057-498-12349	Sequence 12349, A
13	7	3.2	60	6 US-10-203-138A-10571	Sequence 10571, A
14	7	3.2	60	6 US-10-203-138A-10642	Sequence 10642, A
15	7	3.2	60	6 US-10-203-138A-11163	Sequence 11163, A
16	7	3.2	65	6 US-10-203-138A-13258	Sequence 13258, A
17	7	3.2	90	6 US-10-218-140-4872	Sequence 4872, Ap
18	7	3.2	92	1 PCT-US02-32727-16680	Sequence 16680, A
19	7	3.2	92	6 US-10-057-498-16680	Sequence 16680, A
20	7	3.2	98	5 US-09-724-676-55130	Sequence 55130, A
21	7	3.2	98	5 US-09-724-676A-55130	Sequence 55130, A
22	7	3.2	104	6 US-10-264-237-2738	Sequence 2738, Ap
23	7	3.2	111	6 US-10-258-662-47	Sequence 47, Appl
24	7	3.2	113	5 US-09-513-999C-6013	Sequence 6013, Ap
25	7	3.2	117	1 PCT-US02-32727-358	Sequence 358, App
26	7	3.2	117	6 US-10-057-498-358	Sequence 358, App

27	7	3.2	118	5 US-09-724-676-86085	Sequence 86085, A
28	7	3.2	118	5 US-09-724-676A-86085	Sequence 86085, A
29	7	3.2	119	5 US-09-513-999C-8092	Sequence 8092, Ap
30	7	3.2	136	5 US-09-724-676-91178	Sequence 91178, A
31	7	3.2	136	5 US-09-724-676-91179	Sequence 91179, A
32	7	3.2	136	5 US-09-724-676-91190	Sequence 91190, A
33	7	3.2	136	5 US-09-724-676-91191	Sequence 91191, A
34	7	3.2	136	5 US-09-724-676-91192	Sequence 91192, A
35	7	3.2	136	5 US-09-724-676-91193	Sequence 91193, A
36	7	3.2	136	5 US-09-724-676-91194	Sequence 91194, A
37	7	3.2	136	5 US-09-724-676-91201	Sequence 91201, A
38	7	3.2	136	5 US-09-724-676A-91178	Sequence 91178, A
39	7	3.2	136	5 US-09-724-676A-91179	Sequence 91179, A
40	7	3.2	136	5 US-09-724-676A-91190	Sequence 91190, A
41	7	3.2	136	5 US-09-724-676A-91191	Sequence 91191, A
42	7	3.2	136	5 US-09-724-676A-91192	Sequence 91192, A
43	7	3.2	136	5 US-09-724-676A-91193	Sequence 91193, A
44	7	3.2	136	5 US-09-724-676A-91194	Sequence 91194, A
45	7	3.2	136	5 US-09-724-676A-91201	Sequence 91201, A

ALIGNMENTS

RESULT 1
US-10-218-140-1314
; Sequence 1314, Application US/10218140
; GENERAL INFORMATION:
; APPLICANT: Leach, Martin D.
; APPLICANT: Shimkets, Richard A.
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES AND POLYPEPTIDES
; FILE REFERENCE: ENCODED THEREBY
; FILE REFERENCE: 15966-543 CON
; CURRENT APPLICATION NUMBER: US/10/218,140
; CURRENT FILING DATE: 2002-08-12
; PRIOR APPLICATION NUMBER: 09/540,763
; PRIOR FILING DATE: 2000-03-30
; PRIOR APPLICATION NUMBER: 60/127,728
; PRIOR FILING DATE: 1999-04-05
; PRIOR APPLICATION NUMBER: 60/127,636
; PRIOR FILING DATE: 1999-04-02
; PRIOR APPLICATION NUMBER: 60/127,607
; PRIOR FILING DATE: 1999-03-31
; NUMBER OF SEQ ID NOS: 6322
; SOFTWARE: Curanator Version 1.0
; SEQ ID NO 1314
; LENGTH: 121
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-218-140-1314

Query Match 3.7%; Score 8; DB 6; Length 121;
Best Local Similarity 100.0%; Pred. No. 6.7;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 121 AAAPAAP 128

Db 103 AAAPAAP 110

RESULT 2

US-10-337-312-10
; Sequence 10, Application US/10337312
; GENERAL INFORMATION:
; APPLICANT: ALONSO BEDATE, Carlos
; APPLICANT: REQUENA ROLANIA, Jose M.
; APPLICANT: SOTO ALVAREZ, Manuel
; TITLE OF INVENTION: CHIMERIC GENE FORMED OF THE DNA SEQUENCES THAT ENCODE THE ANTI
; TITLE OF INVENTION: DETERMINANTS OF FOUR PROTEINS OF L. INFANTUM, USEFUL FOR SERO
; TITLE OF INVENTION: CANINE LEISHMANIOSIS AND PROTEIN OBTAINED
; FILE REFERENCE: BEDATE-1A
; CURRENT APPLICATION NUMBER: US/10/337,312
; CURRENT FILING DATE: 2003-01-07

```
; PRIOR APPLICATION NUMBER: US/09/788,345
; PRIOR FILING DATE: 2001-02-21
; PRIOR APPLICATION NUMBER: 09/219,306
; PRIOR FILING DATE: 1998-12-23
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 10
; LENGTH: 383
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chimeric
US-10-337-312-10
```

```
Query Match 3.7%; Score 8; DB 6; Length 383;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 120 AAAAPAAP 127
Db 354 AAAAPAAP 361
```

RESULT 3

```
US-10-337-312-12
; Sequence 12, Application US/10337312
; GENERAL INFORMATION:
; APPLICANT: ALONSO BEDATE, Carlos
; APPLICANT: REQUENA ROLANTA, Jose M.
; APPLICANT: SOTO ALVAREZ, Manuel
; TITLE OF INVENTION: CHIMERIC GENE FORMED OF THE DNA SEQUENCES THAT ENCODE THE ANTIGEN
; TITLE OF INVENTION: DETERMINANTS OF FOUR PROTEINS OF L. INFANTUM, USEFUL FOR SEROLOG
; TITLE OF INVENTION: CANINE LEISHMANIOSIS AND PROTEIN OBTAINED
; FILE REFERENCE: BEDATE-1A
; CURRENT APPLICATION NUMBER: US/10/337,312
; CURRENT FILING DATE: 2003-01-07
; PRIOR APPLICATION NUMBER: US/09/788,345
; PRIOR FILING DATE: 2001-02-21
; PRIOR APPLICATION NUMBER: 09/219,306
; PRIOR FILING DATE: 1998-12-23
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 12
; LENGTH: 412
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: recombinant sequence from vectors p031 and pma1
US-10-337-312-12
```

```
Query Match 3.7%; Score 8; DB 6; Length 412;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 120 AAAAPAAP 127
Db 383 AAAAPAAP 390
```

RESULT 4

```
US-10-198-723-14
; Sequence 14, Application US/10198723
; GENERAL INFORMATION:
; APPLICANT: Dang, Van-Dinh
; APPLICANT: Okamuro, Jack
; TITLE OF INVENTION: CHIMERIC POLYPEPTIDES MODULATING
; TITLE OF INVENTION: DEVELOPMENT
; FILE REFERENCE: 11696-043001
; CURRENT APPLICATION NUMBER: US/10/198,723
; CURRENT FILING DATE: 2002-07-05
; PRIOR APPLICATION NUMBER: US 60/303,654
; PRIOR FILING DATE: 2001-07-06
; NUMBER OF SEQ ID NOS: 100
```

```
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 428
; TYPE: PRT
; ORGANISM: Zea mays
US-10-198-723-14
```

```
Query Match 3.7%; Score 8; DB 6; Length 428;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 131 AEPAAAAAP 138
Db 292 AEPAAAAAP 299
```

RESULT 5

```
US-10-203-138A-11635
; Sequence 11635, Application US/102031138A
; GENERAL INFORMATION:
; APPLICANT: Molecular Dynamics, Inc.
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
; TITLE OF INVENTION: ANALYSIS OF GENE EXPRESSION IN BT 474
; FILE REFERENCE: PB 0004 WO 8
; CURRENT APPLICATION NUMBER: US/10/203,138A
; CURRENT FILING DATE: 2002-08-02
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 04 February 2000 (04.02.00)
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 26 May 2000 (26.05.00)
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 03 August 2000 (03.08.00)
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 03 October 2000 (03.10.00)
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 27 September 2000 (27.09.00)
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 21 September 2000 (21.09.00)
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 30 June 2000 (30.06.00)
; NUMBER OF SEQ ID NOS: 15438
; SOFTWARE: Molecular Dynamics Sequence Listing Engine
; SEQ ID NO 11635
; LENGTH: 511
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AB023049.1
; FEATURE:
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 7.2
; FEATURE:
; OTHER INFORMATION: SWISSPROT HIT: Q02817, EVALUE 3.00e-24
; FEATURE:
; OTHER INFORMATION: EST_HUMAN HIT: AA487852.1, EVALUE 2.00e-73
US-10-203-138A-11635
```

```
Query Match 3.7%; Score 8; DB 6; Length 511;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 204 AAPEPHPS 211
Db 95 AAPEPHPS 102
```

RESULT 6

```
US-10-263-929-103
; Sequence 103, Application US/10263929
; GENERAL INFORMATION:
```


; APPLICANT: Kim, Jaeseob
 ; APPLICANT: Galant, Ron
 ; TITLE OF INVENTION: Alzheimer's Disease Linked Genes
 ; FILE REFERENCE: LSD-07417
 ; CURRENT APPLICATION NUMBER: US/10/263,929
 ; CURRENT FILING DATE: 2002-10-03
 ; NUMBER OF SEQ ID NOS: 213
 ; SOFTWARE: PatentIn version 3.2
 ; SEQ ID NO 103
 ; LENGTH: 512
 ; TYPE: PRT
 ; ORGANISM: Drosophila melanogaster
 US-10-263-929-103

Query Match 3.7%; Score 8; DB 6; Length 512;
 Best Local Similarity 100.0%; Pred. No. 24;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 120 AAAAPAAP 127
 Db 173 AAAAPAAP 180

RESULT 7
 US-09-724-676-91213
 ; Sequence 91213, Application US/09724676
 ; GENERAL INFORMATION:
 ; APPLICANT: Compugen LTD
 ; TITLE OF INVENTION: Variants of alternative splicing
 ; FILE REFERENCE: 129181.4 Compugen
 ; CURRENT APPLICATION NUMBER: US/09/724,676
 ; CURRENT FILING DATE: 2000-11-28
 ; NUMBER OF SEQ ID NOS: 97222
 ; SOFTWARE: PatentIn version 3.2
 ; SEQ ID NO 91213
 ; LENGTH: 51
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-09-724-676-91213

Query Match 3.2%; Score 7; DB 5; Length 51;
 Best Local Similarity 100.0%; Pred. No. 27;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 120 AAAAPAA 126
 Db 28 AAAAPAA 34

RESULT 8
 US-09-724-676A-91213
 ; Sequence 91213, Application US/09724676A
 ; GENERAL INFORMATION:
 ; APPLICANT: Compugen LTD
 ; TITLE OF INVENTION: Variants of alternative splicing
 ; FILE REFERENCE: 129181.4 Compugen
 ; CURRENT APPLICATION NUMBER: US/09/724,676A
 ; CURRENT FILING DATE: 2000-11-28
 ; NUMBER OF SEQ ID NOS: 97222
 ; SOFTWARE: PatentIn version 3.2
 ; SEQ ID NO 91213
 ; LENGTH: 51
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-09-724-676A-91213

Query Match 3.2%; Score 7; DB 5; Length 51;
 Best Local Similarity 100.0%; Pred. No. 27;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 120 AAAAPAA 126
 Db 28 AAAAPAA 34

RESULT 9
 PCT-US02-32727-19524
 ; Sequence 19524, Application PC/TUS0232727
 ; GENERAL INFORMATION:
 ; APPLICANT: Mitcham, Jennifer
 ; APPLICANT: Skeiky, Yasir
 ; APPLICANT: Persing, David
 ; APPLICANT: Bhatia, Ajay
 ; APPLICANT: Maisonneuve, Jean Francois
 ; APPLICANT: Zhang, Yanni
 ; APPLICANT: Wang, Siqing
 ; APPLICANT: Jen, Shvian
 ; APPLICANT: Lodes, Michael
 ; APPLICANT: Benson, Darin
 ; APPLICANT: Jones, Robert
 ; APPLICANT: Carter, Darrick
 ; APPLICANT: Barth, Brenda
 ; APPLICANT: Douglass, John
 ; TITLE OF INVENTION: Compositions and Methods for the Therapy and Diagnosis of Acne
 ; FILE REFERENCE: 210121.514C1
 ; CURRENT APPLICATION NUMBER: PCT/US02/32727
 ; CURRENT FILING DATE: 2002-10-11
 ; NUMBER OF SEQ ID NOS: 30992
 ; SEQ ID NO 19524
 ; LENGTH: 53
 ; TYPE: PRT
 ; ORGANISM: Propioni acnes
 PCT-US02-32727-19524

Query Match 3.2%; Score 7; DB 1; Length 53;
 Best Local Similarity 100.0%; Pred. No. 28;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 210 PPSPLEQ 216
 Db 42 PPSPLEQ 48

RESULT 10
 US-10-057-498-19524
 ; Sequence 19524, Application US/10057498
 ; GENERAL INFORMATION:
 ; APPLICANT: Mitcham, Jennifer
 ; APPLICANT: Skeiky, Yasir
 ; APPLICANT: Persing, David
 ; TITLE OF INVENTION: Compositions and Methods for the Therapy and Diagnosis of Acne
 ; FILE REFERENCE: 210121.514
 ; CURRENT APPLICATION NUMBER: US/10/057,498
 ; CURRENT FILING DATE: 2001-04-20
 ; NUMBER OF SEQ ID NOS: 29212
 ; SEQ ID NO 19524
 ; LENGTH: 53
 ; TYPE: PRT
 ; ORGANISM: Propioni acnes
 US-10-057-498-19524

Query Match 3.2%; Score 7; DB 6; Length 53;
 Best Local Similarity 100.0%; Pred. No. 28;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 210 PPSPLEQ 216
 Db 42 PPSPLEQ 48

RESULT 11
 PCT-US02-32727-12349
 ; Sequence 12349, Application PC/TUS0232727
 ; GENERAL INFORMATION:
 ; APPLICANT: Mitcham, Jennifer
 ; APPLICANT: Skeiky, Yasir

```
; APPLICANT: Persing, David
; APPLICANT: Bhatia, Ajay
; APPLICANT: Maisonneuve, Jean Francois
; APPLICANT: Zhang, Yanni
; APPLICANT: Wang, Siqing
; APPLICANT: Jen, Shyian
; APPLICANT: Lodes, Michael
; APPLICANT: Benson, Darin
; APPLICANT: Jones, Robert
; APPLICANT: Carter, Darrick
; APPLICANT: Barth, Brenda
; APPLICANT: Douglass, John
; TITLE OF INVENTION: Compositions and Methods for the Therapy and Diagnosis of Acnes
; FILE REFERENCE: 210121.514C1
; CURRENT APPLICATION NUMBER: PCT/US02/32727
; CURRENT FILING DATE: 2002-10-11
; NUMBER OF SEQ ID NOS: 30992
; SEQ ID NO 12349
; LENGTH: 59
; TYPE: PRT
; ORGANISM: Propionibacterium acnes
PCT-US02-32727-12349

Query Match          3.2%; Score 7; DB 1; Length 59;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 123 APAAPPI 129
Db 50 APAAPPI 56

RESULT 12
US-10-057-498-12349
; Sequence 12349, Application US/10057498
; GENERAL INFORMATION:
; APPLICANT: Mitcham, Jennifer
; APPLICANT: Skeiky, Yasir
; TITLE OF INVENTION: Compositions and Methods for the Therapy and Diagnosis of Acnes
; FILE REFERENCE: 210121.514
; CURRENT APPLICATION NUMBER: US/10/057,498
; CURRENT FILING DATE: 2001-04-20
; NUMBER OF SEQ ID NOS: 29212
; SEQ ID NO 12349
; LENGTH: 59
; TYPE: PRT
; ORGANISM: Propionibacterium acnes
US-10-057-498-12349

Query Match          3.2%; Score 7; DB 6; Length 59;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 123 APAAPPI 129
Db 50 APAAPPI 56

RESULT 13
US-10-203-138A-10571
; Sequence 10571, Application US/10203138A
; GENERAL INFORMATION:
; APPLICANT: Molecular Dynamics, Inc.
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: PB 0004 WO 8
; CURRENT APPLICATION NUMBER: US/10/203,138A
; CURRENT FILING DATE: 2002-08-02
```

```
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 04 February 2000 (04.02.00)
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 26 May 2000 (26.05.00)
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 03 August 2000 (03.08.00)
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 03 October 2000 (03.10.00)
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 27 September 2000 (27.09.00)
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 21 September 2000 (21.09.00)
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 30 June 2000 (30.06.00)
; NUMBER OF SEQ ID NOS: 15438
; SOFTWARE: Molecular Dynamics Sequence Listing Engine
; SEQ ID NO 10571
; LENGTH: 60
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AP000031.1
; FEATURE:
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2.1
; FEATURE:
; OTHER INFORMATION: EST_HUMAN HIT: L25488.1, EVALUE 3.00e+00
; OTHER INFORMATION: SWISSPROT HIT: O95104, EVALUE 9.00e-07
US-10-203-138A-10571

Query Match          3.2%; Score 7; DB 6; Length 60;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 120 AAAAPAA 126
Db 45 AAAAPAA 51

RESULT 14
US-10-203-138A-10642
; Sequence 10642, Application US/10203138A
; GENERAL INFORMATION:
; APPLICANT: Molecular Dynamics, Inc.
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
; FILE REFERENCE: PB 0004 WO 8
; CURRENT APPLICATION NUMBER: US/10/203,138A
; CURRENT FILING DATE: 2002-08-02
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 04 February 2000 (04.02.00)
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 26 May 2000 (26.05.00)
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 03 August 2000 (03.08.00)
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 03 October 2000 (03.10.00)
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 27 September 2000 (27.09.00)
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 21 September 2000 (21.09.00)
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 30 June 2000 (30.06.00)
; NUMBER OF SEQ ID NOS: 15438
; SOFTWARE: Molecular Dynamics Sequence Listing Engine
; SEQ ID NO 10642
; LENGTH: 60
; TYPE: PRT
; ORGANISM: Homo sapiens
```

Job time : 25 secs

```
; FEATURE:
; OTHER INFORMATION: MAP TO AP000213.1
; FEATURE:
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2.2
; FEATURE:
; OTHER INFORMATION: EST_HUMAN HIT: L25488.1, EVALUE 3.00e+00
; FEATURE:
; OTHER INFORMATION: SWISSPROT HIT: O95104, EVALUE 9.00e-07
US-10-203-138A-10642
```

```
Query Match          3.2%; Score 7; DB 6; Length 60;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 120 AAAAPAA 126
      |||||
DB 45 AAAAPAA 51
```

```
RESULT 15
US-10-203-138A-11163
; Sequence 11163, Application US/10203138A
; GENERAL INFORMATION:
; APPLICANT: Molecular Dynamics, Inc.
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: ANALYSIS OF GENE EXPRESSION IN BT 474
; FILE REFERENCE: PB 0004 WO 8
; CURRENT APPLICATION NUMBER: US/10/203,138A
; CURRENT FILING DATE: 2002-08-02
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 04 February 2000 (84.02.00)
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 26 May 2000 (26.05.00)
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 03 August 2000 (03.08.00)
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 03 October 2000 (03.10.00)
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 27 September 2000 (27.09.00)
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 21 September 2000 (21.09.00)
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 30 June 2000 (30.06.00)
; NUMBER OF SEQ ID NOS: 15438
; SOFTWARE: Molecular Dynamics Sequence Listing Engine
; SEQ ID NO 11163
; LENGTH: 60
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AP000135.1
; FEATURE:
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 3.4
; FEATURE:
; OTHER INFORMATION: EST_HUMAN HIT: L25488.1, EVALUE 3.00e+00
; FEATURE:
; OTHER INFORMATION: SWISSPROT HIT: O95104, EVALUE 9.00e-07
US-10-203-138A-11163
```

```
Query Match          3.2%; Score 7; DB 6; Length 60;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 120 AAAAPAA 126
      |||||
DB 45 AAAAPAA 51
```

Search completed: February 12, 2003, 14:03:20

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2003, 13:55:40 ; Search time 21 seconds
(without alignments)
1002.545 Million cell updates/sec

Title: US-09-923-236-2

Perfect score: 219

Sequence: 1 MKLLMWACIVCFARRRF.....TAKPAPEPHPSLEQANQ 219

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 283224 seqs, 96134422 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : PIR_73.*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	9	4.1	581	2 T36267	probable glutamyl-nitrate reductase
2	9	4.1	864	1 JC1422	nitrate reductase
3	8	3.7	110	1 R6BYP3	60s acidic ribosomal L7/L12 ribosomal p
4	8	3.7	129	2 H86500	ribosomal protein
5	8	3.7	129	2 C72122	probable lipid tra
6	8	3.7	193	2 H86404	cytochrome c, memb
7	8	3.7	228	2 E87612	hypothetical prote
8	8	3.7	268	2 G70645	probable ABC trans
9	8	3.7	276	2 G95887	DNA binding protei
10	8	3.7	291	2 T02434	electron transfer
11	8	3.7	325	2 T48873	TPR domain protein
12	8	3.7	336	2 A87300	pupal cuticula pro
13	8	3.7	353	2 S41958	phenoxymethyl ester
14	8	3.7	381	2 T46827	cellular tumor ant
15	8	3.7	396	1 JH0633	probable farnesyl
16	8	3.7	495	2 T52066	transcription enha
17	8	3.7	506	2 A40679	transcription enha
18	8	3.7	523	2 B40679	hypothetical prote
19	8	3.7	550	2 C75557	H+-transporting tw
20	8	3.7	574	2 S23530	probable cell divi
21	8	3.7	611	2 C96030	GTP-binding regula
22	8	3.7	846	2 S52418	kelch protein, lon
23	8	3.7	1476	2 A45773	hypothetical prote
24	8	3.7	4307	2 T20721	photosystem I chai
25	7	3.2	38	2 S23173	hypothetical prote
26	7	3.2	78	2 E33446	antifreeze protein
27	7	3.2	82	1 PFPLAW	antifreeze protein
28	7	3.2	82	2 S02326	antifreeze protein
29	7	3.2	82	2 A05161	antifreeze protein

30 7 3.2 82 2 I51125 antifreeze protein
31 7 3.2 100 2 F95852 hypothetical prote
32 7 3.2 101 2 H69091 ribosomal protein
33 7 3.2 102 2 D64363 ribosomal protein
34 7 3.2 106 1 R6DOP2 acidic ribosomal p
35 7 3.2 107 1 R6KMIC acidic ribosomal p
36 7 3.2 109 1 R6UTP1 acidic ribosomal p
37 7 3.2 110 2 T37490 ribosomal protein
38 7 3.2 111 2 E72524 probable ribosomal
39 7 3.2 111 2 B95857 hypothetical prote
40 7 3.2 112 1 R5FFZE acidic ribosomal p
41 7 3.2 112 2 S54179 acidic ribosomal p
42 7 3.2 112 2 E86141 protein T25K16.9 l
43 7 3.2 114 1 R5RT12 acidic ribosomal p
44 7 3.2 114 1 R5HS2H ribosomal protein
45 7 3.2 114 2 F84266 50S ribosomal prot

ALIGNMENTS

RESULT 1

T36267
probable glutamyl-tRNA reductase - Streptomyces coelicolor
C:Species: Streptomyces coelicolor
C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C:Accession: T36267
R:Murphy, L.; Harris, D.; James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, June 1999
A:Reference number: Z21576
A:Accession: T36267
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-581 <MUR>
A:Cross-references: EMBL:AL079345; PIDN:CAB45353.1; GSPDB:GN00070; SCOEDB:SCE68.17c
A:Experimental source: strain A3(2)
C:Genetics:
A:Gene: SCOEDB:SCE68.17c

Query Match 4.1% Score 9; DB 2; Length 581;
Best Local Similarity 100.0%; Pred.No. 2;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 154 VAAEPAAEA 162

Db 36 VAAEPAAEA 44

RESULT 2

JC1422
nitrate reductase (NADH) (EC 1.7.1.1) - Volvox carteri
C:Species: Volvox carteri
C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 03-Jun-2002
C:Accession: JC1422; S22192
R:Gruber, H.; Goetlinc, S.D.; Kirk, D.L.; Schmitt, R.
Gene 120, 75-83, 1992
A:Title: The nitrate reductase-encoding gene of Volvox carteri: Map location, sequenc
A:Reference number: JC1422; MUID:93013022; PMID:1398126
A:Accession: JC1422
A:Molecule type: DNA
A:Residues: 1-864 <GR2>
A:Cross-references: EMBL:X64136; NID:g21993; PIDN:CAA45497.1; PTD:g21994
A:Note: submitted to the EMBL Data Library, January 1992
C:Genetics:
A:Gene: nita

A:Map position: linkage group IX
A:Introns: 183/3; 234/2; 295/1; 339/1; 372/1; 425/2; 521/3; 593/3; 677/3; 797/2
C:Function:

A:Description: catalyzes the reduction of nitrate to nitrite
C:Superfamily: nitrate reductase (NADH); cytochrome b5 core homology; cytochrome-b5 r
C:Keywords: dimer; electron transfer; FAD; flavoprotein; heme; iron; metalloprotein;
E;36-425/Domain: molybdopter-in-binding domain homology <PCO>
F;497-571/Domain: cytochrome b5 core homology <CB5>

F:613-864/Domain: cytochrome-b5 reductase homology <CBR>
F:139/Binding site: molybdopterin (Cys) (covalent) #status predicted
F:532,555/Binding site: heme iron (His) (axial ligands) #status predicted

Query Match 4.1%; Score 9; DB 1; Length 864;
Best Local Similarity 100.0%; Pred. No. 2.8;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 120 AAAAPAAPP 128

|||||

Db 478 AAAAPAAPP 486

RESULT 3

RBYP3

60s acidic ribosomal protein p1-alpha - fission yeast (Schizosaccharomyces pombe)

N:Alternate names: acidic ribosomal protein 3; ribosomal protein YL12eII

C:Species: Schizosaccharomyces pombe

C>Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 10-Dec-1999

C:Accession: C34715; T40352

R:Beltrame, M.; Bianchi, M.E.

Mol. Cell. Biol. 10, 2341-2348, 1990

A:Title: A gene family for acidic ribosomal proteins in Schizosaccharomyces pombe: two e

A:Reference number: A34715; MUID:90220620; PMID:2325655

A:Accession: C34715

A:Molecule type: DNA

A:Residues: 1-110 <BEL>

A:Cross-references: EMBL:M3139; NID:g173467; PIDN:AAA35336.1; PID:g173468

A:Wood, V.; Rajendrem, M.A.; Barrell, B.G.; Skelton, J.; Churcher, C.M.

submitted to the EMBL Data Library, March 1997

A:Reference number: Z21922

A:Accession: T40352

A>Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-110 <WOO>

A:Cross-references: EMBL:AL022070; PIDN:CAAL7793.1; GSPDB:GN00067; SPDB:SPBC3B9.13c

A:Experimental source: strain 972h-; cosmid c3B9

C:Genetics:

A:Gene: rpa3

A:Map position: 2

A:Introns: 23/3; 103/3

C:Superfamily: rat acidic ribosomal protein P1

C:Keywords: phosphoprotein; protein biosynthesis; ribosome

Query Match 3.7%; Score 8; DB 1; Length 110;

Best Local Similarity 100.0%; Pred. No. 4.4;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 119 SAAAAPAA 126

|||||

Db 64 SAAAAPAA 71

RESULT 4

H86500

L7/L12 ribosomal protein [imported] - Chlamydomophila pneumoniae (strain J138)

C:Species: Chlamydomophila pneumoniae, Chlamydia pneumoniae

C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 23-Mar-2001

C:Accession: H86500

R:Shirai, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.; Is

Nucleic Acids Res. 28, 2311-2314, 2000

A:Title: Comparison of whole genome sequences of chlamydia pneumoniae J138.

A:Reference number: A86491; MUID:20330349; PMID:10871362

A:Accession: H86500

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-129 <STO>

A:Cross-references: GB:BA000008; NID:g8978453; PIDN:BAA98290.1; GSPDB:GN00142

A:Experimental source: strain J138

C:Genetics:

A:Gene: r17

C:Superfamily: Escherichia coli ribosomal protein L12

Query Match 3.7%; Score 8; DB 2; Length 129;
Best Local Similarity 100.0%; Pred. No. 5;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 171 EAPVAAEP 178

|||||

Db 50 EAPVAAEP 57

RESULT 5

C72122

ribosomal protein L7/L12 CP0695 [imported] - Chlamydomophila pneumoniae (strains CWL029

C:Species: Chlamydomophila pneumoniae, Chlamydia pneumoniae

C>Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-May-2000

C:Accession: C72122; G81548

R:Kalmán, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood,

Nature Genet. 21, 385-389, 1999

A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.

A:Reference number: A72000; MUID:99206606; PMID:10192388

A:Accession: C72122

A:Molecule type: DNA

A:Residues: 1-129 <ARN>

A:Cross-references: GB:AE001593; GB:AE001363; NID:g4376334; PIDN:AAID18233.1; PID:g437

A:Experimental source: strain CWL029

R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hicke

C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzbe

Nucleic Acids Res. 28, 1397-1406, 2000

A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39

A:Reference number: A81500; MUID:20150255; PMID:10684935

A:Accession: G81548

A:Molecule type: DNA

A:Residues: 1-129 <REA>

A:Cross-references: GB:AE002228; GB:AE002161; NID:g7189606; PIDN:AAF38503.1; PID:g718

A:Experimental source: strain AR39, HL cells

C:Genetics:

A:Gene: r17; CP0695

C:Superfamily: Escherichia coli ribosomal protein L12

Query Match 3.7%; Score 8; DB 2; Length 129;

Best Local Similarity 100.0%; Pred. No. 5;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 171 EAPVAAEP 178

|||||

Db 50 EAPVAAEP 57

RESULT 6

H86404

Probable lipid transfer protein [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001

C:Accession: H86404

R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon

Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, T.H.; Dewar,

ansen, N.F.; Hughes, B.; Huizar, L.

Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzia

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; MUID:21016719; PMID:11130712

A:Accession: H86404

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-193 <STO>

A:Cross-references: GB:AE005172; NID:g11024871; PIDN:AAG26955.1; GSPDB:GN00141

C:Genetics:

A:Map position: 1

Query Match 3.7%; Score 8; DB 2; Length 193;

Best Local Similarity 100.0%; Pred. No. 7;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 120 AAAAPAAP 127
|||||
Db 20 AAAAPAAP 27

RESULT 7
EB87612
cytochrome c, membrane-bound [imported] - Caulobacter crescentus
C:Species: Caulobacter crescentus
C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C:Accession: E87612
R:Niernman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
B.; Laub, M.T.; DeBOY, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A:Title: Complete Genome Sequence of *Caulobacter crescentus*.
A:Reference number: A87249; MUID:21173698; PMID:11259647
A:Accession: E87612
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-228 <STO>
A:Cross-references: GB:AF005673; NID:g13424561; PIDN:AAK24897.1; GSPDB:GN00148
C:Genetics:
A:Gene: CC2935

Query Match 3.7%; Score 8; DB 2; Length 228;
Best Local Similarity 100.0%; Pred. No. 8;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 120 AAAAPAAP 127
|||||
Db 219 AAAAPAAP 226

RESULT 8
G70645
hypothetical protein RV3134c - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: G70645
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.
Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: G70645
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-268 <COL>
A:Cross-references: GB:283867; GB:AL123456; NID:g3261695; PIDN:CAB06280.1; PID:e290951;
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: RV3134c

Query Match 3.7%; Score 8; DB 2; Length 268;
Best Local Similarity 100.0%; Pred. No. 9.1;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 193 PSPAEPAT 200
|||||
Db 138 PSPAEPAT 145

RESULT 9
G95887
probable ABC transporter permease protein Smb20381 [imported] - *Sinorhizobium meliloti* (C:Species: *Sinorhizobium meliloti*)
C:Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 01-Mar-2002

C:Accession: G95887
R:Finan, T.M.; Weidner, S.; Wong, K.; Buhrmester, J.; Chain, P.; Vorholter, F.J.; Her
Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001
A:Title: The complete sequence of the 1,683-kb pSymB megaplasmid from the N2-fixing e
A:Reference number: A95842; MUID:21396508; PMID:11481431
A:Accession: G95887
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-276 <KUR>
A:Cross-references: GB:AL591985; PIDN:CAC48767.1; PID:g15140240; GSPDB:GN00167
A:Experimental source: strain 1021, megaplasmid pSymB
R:Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubl
pella, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.
L.; Hyman, R.W.; Jones, T.
Science 293, 668-672, 2001
A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelau
heault, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh,
A:Title: The composite genome of the legume symbiont *Sinorhizobium meliloti*.
A:Reference number: A96039; MUID:21368234; PMID:11474104
A:Contents: annotation
C:Genetics:
A:Gene: Smb20381
A:Genome: plasmid
C:Superfamily: spermidine/putrescine transport system permease protein potI

Query Match 3.7%; Score 8; DB 2; Length 276;
Best Local Similarity 100.0%; Pred. No. 9.3;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 130 AAEPAAAA 137
|||||
Db 268 AAEPAAAA 275

RESULT 10
T02434
DNA binding protein EREBP-4 - common tobacco
C:Species: *Nicotiana tabacum* (common tobacco)
C:Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 31-Mar-2001
C:Accession: T02434
R:Ohme-Takagi, M.; Shinshi, H.
Plant Cell 7, 173-182, 1995
A:Title: Ethylene-inducible DNA binding proteins that interact with an ethylene respo
A:Reference number: Z14671; MUID:95276459; PMID:7756828
A:Accession: T02434
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-291 <OHM>
A:Cross-references: EMBL:D38125; NID:g790361; PIDN:BAA07323.1; PID:g1208497
A:Experimental source: strain BV4; tissue-type leaf
C:Superfamily: Arabidopsis thaliana hypothetical protein T19P19.170

Query Match 3.7%; Score 8; DB 2; Length 291;
Best Local Similarity 100.0%; Pred. No. 9.8;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 133 PAAAPLPT 140
|||||
Db 247 PAAAPLPT 254

RESULT 11
T48873
electron transfer protein [imported] - *Frateuria* sp. (strain ANA-18)
C:Species: *Frateuria* sp.
A:Variety: strain ANA-18
C:Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 09-Jun-2000
C:Accession: T48873
R:Murakami, S.; Takashima, A.; Takemoto, J.; Takenaka, S.; Shinke, R.; Aoki, K.
Gene 226, 189-198, 1999
A:Title: Cloning and sequence analysis of two catechol-degrading gene clusters from t
A:Reference number: Z24833
A:Accession: T48873

A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-325 <MUR>
 A:Cross-references: EMBL:AB009373; PTDN:BA075213.1
 A:Experimental source: strain ANA-18
 C:Superfamily: phthalate dioxygenase reductase; cytochrome-b5 reductase homology; ferredoxin

Query Match 3.7%; Score 8; DB 2; Length 325;
 Best Local Similarity 100.0%; Pred. No. 11; Mismatches 0; Indels 0; Gaps 0;
 Matches 8; Conservative 0

Qy 130 AAEPAAAA 137
 |||||
 Db 223 AAEPAAAA 230

RESULT 12

A87300
 TPR domain protein [imported] - Caulobacter crescentus
 C:Species: Caulobacter crescentus
 C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
 C:Accession: A87300
 R:Neiman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.; B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolonitskii, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.; Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
 A:Title: Complete Genome Sequence of Caulobacter crescentus.
 A:Reference number: A87249; MUID:21173698; PMID:11259647
 A:Accession: A87300
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-336 <STO>
 A:Cross-references: GB:AE005673; NID:gl3421571; PTDN:AAK22397.1; GSPDB:GN001448
 C:Genetics: CC0410
 A:Gene: CC0410

Query Match 3.7%; Score 8; DB 2; Length 336;
 Best Local Similarity 100.0%; Pred. No. 11; Mismatches 0; Indels 0; Gaps 0;
 Matches 8; Conservative 0

Qy 120 AAAPAAP 127
 |||||
 Db 112 AAAPAAP 119

RESULT 13

S41958
 pupal cuticle protein precursor - greater wax moth
 C:Species: Galleria mellonella (greater wax moth)
 C:Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 19-May-2000
 C:Accession: S41958
 R:Kollberg, U.; Obermeier, B.; Hirsch, H.; Kelber, G.; Wolbert, P.
 submitted to the EMBL Data Library, February 1994
 A:Description: Expression cloning and characterization of a pupal cuticle protein cDNA
 A:Reference number: S41958
 A:Accession: S41958
 A:Molecule type: mRNA
 A:Residues: 1-353 <KOL>
 A:Cross-references: EMBL:X77514; NID:g453403; PTDN:CAA54650.1; PID:g453404
 F:1-15/Domain: signal sequence #status predicted <SIG>
 F:16-353/Product: pupal cuticle protein #status predicted <MAT>

Query Match 3.7%; Score 8; DB 2; Length 353;
 Best Local Similarity 100.0%; Pred. No. 11; Mismatches 0; Indels 0; Gaps 0;
 Matches 8; Conservative 0

Qy 119 SAAAPAA 126
 |||||
 Db 236 SAAAPAA 243

RESULT 14

T46827

phenoxybenzoate dioxygenase [imported] - Acinetobacter lwoffii
 C:Species: Acinetobacter lwoffii
 C:Date: 17-Mar-2000 #sequence_revision 17-Mar-2000 #text_change 02-Sep-2000
 C:Accession: T46827
 R:Kim, S.I.; Leem, S.H.; Choi, J.S.; Chung, Y.H.; Kim, S.; Park, Y.M.; Ha, K.S.
 submitted to the EMBL Data Library, November 1996
 A:Reference number: Z24100
 A:Accession: T46827
 A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA
 A:Residues: 1-381 <KIM>
 A:Cross-references: EMBL:U77659; PTDN:AAC31770.1
 A:Experimental source: strain K24
 C:Superfamily: phthalate dioxygenase reductase; cytochrome-b5 reductase homology; ferredoxin

Query Match 3.7%; Score 8; DB 2; Length 381;
 Best Local Similarity 100.0%; Pred. No. 12; Mismatches 0; Indels 0; Gaps 0;
 Matches 8; Conservative 0

Qy 130 AAEPAAAA 137
 |||||
 Db 279 AAEPAAAA 286

RESULT 15

JH0633
 cellular tumor antigen p53 - golden hamster
 N:Alternate names: tumor-suppressor protein p53
 C:Species: Mesocricetus auratus (golden hamster)
 C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
 C:Accession: JH0633
 R:Legros, Y.; McIntyre, P.; Soussi, T.
 Gene 112, 247-250, 1992
 A:Title: The cDNA cloning and immunological characterization of hamster p53.
 A:Reference number: JH0633; MUID:92210007; PMID:1555773
 A:Accession: JH0633
 A:Molecule type: mRNA
 A:Residues: 1-396 <LEG>
 A:Cross-references: GB:M75144; NID:gl91414; PTDN:AAA37085.1; PID:gl91415
 A:Experimental source: kidney, strain MP1
 C:Genetics: C

A:Gene: p53
 C:Superfamily: cellular tumor antigen p53
 C:Keywords: apoptosis; cell division control; DNA binding; homotrimer; nucleus; phosphorylation; zinc (Cys, His, Cys, Cys) #status predicted
 F:179,182,241,245/Binding site: zinc (Cys, His, Cys, Cys) #status predicted
 F:395/Binding site: phosphoryl-RNA (Ser) (covalent) #status predicted

Query Match 3.7%; Score 8; DB 1; Length 396;
 Best Local Similarity 100.0%; Pred. No. 13; Mismatches 0; Indels 0; Gaps 0;
 Matches 8; Conservative 0

Qy 120 AAAAPAAP 127
 |||||
 Db 66 AAAAPAAP 73

Search completed: February 12, 2003, 13:59:37
 Job time : 23 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2003, 13:49:34 ; Search time 16 Seconds
(without alignments)
567.707 Million cell updates/sec

Title: US-09-923-236-2

Perfect score: 219
Sequence: 1 MKLLWACIVCAFARKRF.....TAKPAAPHPSPSLEQANQ 219

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 112892 seqs, 41476328 residues

Word size : 0

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	9	4.1	322	RLA0_LEICH	P39096 leishmania
2	9	4.1	323	RLA0_LEIIN	P39097 leishmania
3	9	4.1	581	HEMI_STRCO	Q9wx15 streptomyces
4	9	4.1	864	NIA_VOLCA	P36841 volvox cart
5	8	3.7	107	RLA1_LEIPE	O46313 leishmania
6	8	3.7	110	RLA3_SCHPO	P17477 schizosacch
7	8	3.7	128	RL7_CHLPN	Q9z9a1 chlamydia p
8	8	3.7	353	CUP5_GALME	Q24998 gallieria me
9	8	3.7	396	P53_MESAU	Q00366 mesocricetu
10	8	3.7	574	ATP2_CHLRE	P38482 chlamydomon
11	8	3.7	575	MSPA_TREMA	Q9z413 treponema m
12	8	3.7	576	PCR_HELVI	O18473 heliothis v
13	8	3.7	1477	KELC_DROME	Q04652 drosophila
14	8	3.7	2470	NTC2_MOUSE	Q35516 mus musculus
15	7	3.2	82	ANPA_PSEAM	P04002 pseudopleur
16	7	3.2	101	RL12_METH	P05394 methanobact
17	7	3.2	101	RL12_METHL	O52706 methanococc
18	7	3.2	102	RL12_METJA	P54048 methanococc
19	7	3.2	105	RLA2_DICDI	P22683 dictyosteli
20	7	3.2	107	RLA1_CHLRE	P29763 chlamydomon
21	7	3.2	109	RLA1_TRYCR	P26643 trypanosoma
22	7	3.2	110	RLA2_CRYST	O61463 cryptochito
23	7	3.2	111	RL2A_MAIZE	P46252 zea mays (m
24	7	3.2	111	RL12_AERPE	Q9y9w9 aeropyrum p
25	7	3.2	111	RLA1_CAEEL	P91913 caenorhabd
26	7	3.2	111	RLA2_ASFPU	Q9uuz6 aspergillus
27	7	3.2	112	RLA1_DROME	P08570 drosophila
28	7	3.2	114	RL12_HALN1	P05768 halobacteri
29	7	3.2	114	RLA1_RAT	P19944 rattus norv
30	7	3.2	114	RLA2_EIMTE	Q967v9 eimeria ten
31	7	3.2	120	CUI9_ARADI	P80515 araneus dia
32	7	3.2	125	PSAE_SPIOL	P12354 spinacia ol
33	7	3.2	129	DYLA_CHLRE	Q39591 chlamydomon

```
34 7 3.2 136 1 RL28_HUMAN P46779 homo sapien
35 7 3.2 136 1 RL28_MOUSE P41105 mus musculus
36 7 3.2 136 1 SR14_HUMAN P37108 homo sapien
37 7 3.2 156 1 BCCP_PSEAE P37799 pseudomonas
38 7 3.2 163 1 ATPD_CAEEL O09544 caenorhabdi
39 7 3.2 168 1 ATPD_HUMAN P30049 homo sapien
40 7 3.2 199 1 CYCY_RHOCA Q05389 rhodobacter
41 7 3.2 208 1 RS6_MYCGE P47336 mycoplasma
42 7 3.2 215 1 EFID_WHEAT P29546 triticum ae
43 7 3.2 215 1 RS6_MYCPN P75543 mycoplasma
44 7 3.2 220 1 COAT_CMVSI Q00467 cymbidium m
45 7 3.2 222 1 EFID_ORYSA P29545 oryza sativ
```

ALIGNMENTS

```
RESULT 1
ID RLA0_LEICH STANDARD; PRT; 322 AA.
AC P39096;
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE 60S acidic ribosomal protein P0.
GN LCP0.
OS Leishmania chagasi.
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=44271;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ISOLATE MHOM/BR/82/BA-2;
RX MEDLINE=94222525; Pubmed=7513304;
RA Skeiky Y.A.W., Benson D.R., Elwasila M., Badaro R., Burns J.M. Jr.,
RA Reed S.G.;
RT "Antigens shared by Leishmania species and Trypanosoma cruzi:
RT immunological comparison of the acidic ribosomal P0 proteins.";
RL Infect. Immun. 62:1643-1651(1994).
CC -!- FUNCTION: RIBOSOMAL PROTEIN P0 IS THE FUNCTIONAL EQUIVALENT
CC OF E. COLI PROTEIN L10.
CC -!- SUBUNIT: P0 FORMS A PENTAMERIC COMPLEX BY INTERACTION WITH
CC DIMERS OF P1 AND P2.
CC -!- SIMILARITY: BELONGS TO THE L10P FAMILY OF RIBOSOMAL PROTEINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC EMBL; L29300; AAA29263.1;
CC InterPro; IPR001813; 60S_Ribosomal.
CC InterPro; IPR001790; Ribosomal_L10.
CC Pfam; PF00428; 60S_Ribosomal; 1.
CC Pfam; PF00466; Ribosomal_L10; 1.
CC Ribosomal protein; Phosphorylation.
SQ SEQUENCE 322 AA; 34594 MW; 2389F125356D26D2 CRC64;
```

Query Match 4.1%; Score 9; DB 1; Length 322;
Best Local Similarity 100.0%; Pred. No. 0.56;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 130 AAEPAAAAAP 138
Db 289 AAEPAAAAAP 297
|||||
```

```
RESULT 2
RLA0_LEIIN STANDARD; PRT; 323 AA.
ID RLA0_LEIIN
AC P39097;
```

```

DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE 60S acidic ribosomal protein P0.
DE LIPO-A AND LIPO-B.
GN Leishmania infantum.
OS Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=5671;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LEM 75;
RX MEDLINE=9408674; PubMed=8264730;
RA Soto M., Requena J.M., Alonso C.;
RT "Isolation, characterization and analysis of the expression of the
RT Leishmania ribosomal P0 protein genes.";
RL Mol. Biochem. Parasitol. 61:265-274(1993).
CC -!- FUNCTION: RIBOSOMAL PROTEIN P0 IS THE FUNCTIONAL EQUIVALENT
CC OF E.COLI PROTEIN L10.
CC -!- SUBUNIT: P0 FORMS A PENTAMERIC COMPLEX BY INTERACTION WITH
CC DIMERS OF P1 AND P2.
CC -!- SIMILARITY: BELONGS TO THE L10P FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X72714; CAA51264.1; -.
DR EMBL: X72714; CAA51263.1; -.
DR InterPro: IPR001813; 60S_ribosomal.
DR Pfam: PF00428; 60S_ribosomal; 1.
DR Pfam: PF00466; Ribosomal_L10; 1.
KW Ribosomal protein; Phosphorylation.
SQ SEQUENCE 323 AA; 34771 MW; 14967BD7A439D69E CRC64;

Query Match 4.1%; Score 9; DB 1; Length 323;
Best Local Similarity 100.0%; Pred. No. 0.57;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 130 AAEPAAAAP 138
DB 290 AAEPAAAAP 298
|||||

RESULT 3
HEM1_STRCO STANDARD; PRT; 581 AA.
AC Q9WX15;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Glutamyl-tRNA reductase (EC 1.2.1.-) (GluTR).
GN HEMA OR SC03319 OR SCE68.17C.
OS Streptomyces coelicolor.
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RX MEDLINE=21956410; PubMed=12000953;
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA Rabinowitz E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares S., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J., Barrall B.G., Parkhill J.,

```

```

RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RL coelicolor A3(2).";
CC -!- CATALYTIC ACTIVITY: Glutamyl-tRNA(Glu) + NADPH = glutamate-1-
CC semialdehyde + NADP(+) + tRNA(Glu).
CC -!- PATHWAY: Porphyrin biosynthesis by the C5 pathway; first step.
CC -!- SIMILARITY: BELONGS TO THE GLUTAMYL-TRNA REDUCTASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AL079345; CAB45353.1; -.
DR InterPro: IPR000343; GluTR.
DR Pfam: PF00745; GluTR; 1.
DR TIGRFAMs: TIGR01035; hemaA; 1.
DR PROSITE: PS00747; GLUTR; 1.
KW Porphyrin biosynthesis; Oxidoreductase; NADP; Complete proteome.
FT DOMAIN 292 416 INSERT.
FT ACT_SITE 50 50 NUCLEOPHILE (BY SIMILARITY).
FT ACT_SITE 99 99 BASE (BY SIMILARITY).
SQ SEQUENCE 581 AA; 60562 MW; D4E256B105AFA037 CRC64;

Query Match 4.1%; Score 9; DB 1; Length 581;
Best Local Similarity 100.0%; Pred. No. 0.93;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 154 VAAEPRAEA 162
DB 36 VAAEPRAEA 44
|||||

RESULT 4
NIA_VOLCA STANDARD; PRT; 864 AA.
ID NIA_VOLCA
AC P36841;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DE Nitrate reductase (EC 1.6.6.1) (NR).
GN NITA.
OS Volvox carteri.
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC Volvocaceae; Volvox.
OX NCBI_TaxID=3067;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=f. Nagariensis / HK10;
RX MEDLINE=93013022; PubMed=1398126;
RA Gruber H., Goetinck S.D., Kirk D.L., Schmitt R.;
RT "The nitrate reductase-encoding gene of Volvox carteri: map location,
RT sequence and induction kinetics.";
RL Gene 120:75-83(1992).
CC -!- FUNCTION: NITRATE REDUCTASE IS A KEY ENZYME INVOLVED IN THE FIRST
CC STEP OF NITRATE ASSIMILATION IN PLANTS, FUNGI AND BACTERIA.
CC -!- CATALYTIC ACTIVITY: NADH + nitrate = NAD(+) + nitrite + H(2)O.
CC -!- COFACTOR: EACH SUBUNIT OF THE ENZYME CONTAINS 1 EQUIVALENT OF FAD,
CC HEME IRON, AND MOLYBDENUM-PTERIN AS PROSTHETIC GROUPS. THE HEME
CC GROUP IS CALLED CYTOCHROME B-557.
CC -!- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -!- INDUCTION: BY NITRATE.
CC -!- SIMILARITY: TO EUKARYOTIC MOLYBDOPTERIN OXIDOREDUCTASES IN THE
CC N-TERMINAL DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 CYTOCHROME B5 FAMILY, HEME-BINDING DOMAIN.
CC -!- SIMILARITY: TO FAD/NAD-BINDING CYTOCHROME REDUCTASES IN THE
CC C-TERMINAL DOMAIN.
CC -----

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

DR EMBL: X64136; CAA45497.1; -
DR PIR: JC1422; JC1422.
DR PIR: S22192; S22192.
DR HSSP: P04166; 1B5M.
DR InterPro: IPR001199; Cyt_B5.
DR InterPro: IPR001834; Cyt_B5_reductase.
DR InterPro: IPR000572; Euk_Mb_oxred.
DR InterPro: IPR005066; Mo-co_dimer.
DR InterPro: IPR001433; Oxred_FAD/NAD(P).
DR Pfam: PF00173; heme_1; 1.
DR Pfam: PF00174; oxidored_molyb; 1.
DR Pfam: PF00175; NAD_binding; 1.
DR Pfam: PF00970; FAD_binding_6; 1.
DR Pfam: PF03404; Mo-co_dimer; 1.
DR PRINTS: PR00406; CYTB5RDTASE.
DR PRINTS: PR00363; CYTOCHROME B5.
DR PRINTS: PR00407; EUOOPTERIN.
DR PRODOM: PD000612; Cyt_B5; 1.
DR PROSITE: PS00191; CYTOCHROME_B5_1; 1.
DR PROSITE: PS02025; CYTOCHROME_B5_2; 1.
DR PROSITE: PS00559; MOLYBDOPTERIN_EUK; 1.
KW Oxidoreductase; Flavoprotein; FAD; NAD; Heme; Molybdenum;
KW Nitrate assimilation.
FT METAL 139 139 MOLYBDENUM-PTERIN (POTENTIAL).
FT METAL 193 193 MOLYBDENUM-PTERIN (POTENTIAL).
FT DISULFID 376 376 INTERCHAIN (POTENTIAL).
FT BINDING 532 532 HEME LIGAND (BY SIMILARITY).
FT BINDING 555 555 HEME LIGAND (BY SIMILARITY).
SQ SEQUENCE 864 AA; 96402 MW; 499529652CDDDC17 CRC64;

Query Match 4.1%; Score 9; DB 1; Length 864;
Best Local Similarity 100.0%; Pred. No. 1.3;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 120 AAAAAPAPP 128
Db 478 AAAAAPAPP 486
|||||

RESULT 5
ID RLAL_LEIPE STANDARD; PRT; 107 AA.
AC O46313;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 13-DEC-1998 (Rel. 37, Last annotation update)
DE 60S acidic ribosomal protein P1.
OS Leishmania peruviana.
OC Eukaryota; Eulenzozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=5681;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MHOM/PE/84/LC26;
RA de Los Santos M., Carrillo C., Panebra A., Montoya Y.;
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: PLAYS AN IMPORTANT ROLE IN THE ELONGATION STEP OF
CC PROTEIN SYNTHESIS.
CC -!- SUBUNIT: P1 AND P2 EXIST AS DIMERS AT THE LARGE RIBOSOMAL
CC SUBUNIT.
CC -!- SIMILARITY: BELONGS TO THE L12P FAMILY OF RIBOSOMAL PROTEINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

DR EMBL: AF045249; AAC02701.1; -
DR InterPro: IPR001813; 60S_ribosomal.
DR Pfam: PF00428; 60S_ribosomal; 1.
KW Ribosomal protein.
SQ SEQUENCE 107 AA; 10896 MW; 9117DEC5E490A071 CRC64;

Query Match 3.7%; Score 8; DB 1; Length 107;
Best Local Similarity 100.0%; Pred. No. 1.9;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 119 SAAAAPAA 126
Db 71 SAAAAPAA 78
|||||

RESULT 6
ID RLAA3_SCHPO STANDARD; PRT; 110 AA.
AC P17477;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE 60S acidic ribosomal protein P1-alpha 3 (A3).
GN RPA3 OR SPBC3B9.13C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90220620; PubMed=2325655;
RA Beltrame M., Bianchi M.E.;
RT "A gene family for acidic ribosomal proteins in Schizosaccharomycetes
RT pombe: two essential and two nonessential genes.";
RL Mol. Cell. Biol. 10:2341-2348(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neil S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volckaert G., Aert R., Robben J., Grymoprez B.,
RA Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Goffreau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakowski G.V., Ussery D., Barrell B.G., Nurse P.;
RL "The genome sequence of Schizosaccharomycetes pombe.";
RT Nature 415:871-880(2002).

CC -!- FUNCTION: PLAYS AN IMPORTANT ROLE IN THE ELONGATION STEP OF
CC PROTEIN SYNTHESIS.
CC -!- SUBUNIT: P1 AND P2 EXIST AS DIMERS AT THE LARGE RIBOSOMAL

CC SUBUNIT.
 CC -!- MISCELLANEOUS: YEASTS CONTAIN 4 INDIVIDUAL SMALL RIBOSOMAL A
 CC PROTEINS (RPA) WHICH CAN BE CLASSIFIED INTO TWO COUPLES OF SIMILAR
 CC BUT NOT IDENTICAL SEQUENCES. EACH COUPLE IS DISTINCTLY RELATED TO
 CC ONE OF THE TWO A PROTEINS PRESENT IN MULTICELLULAR ORGANISMS.
 CC -!- MISCELLANEOUS: RPA3 AND RPA4 ARE ESSENTIAL FOR CELL SURVIVAL,
 CC WHEREAS RPA1 AND RPA2 ARE NOT.
 CC -!- SIMILARITY: BELONGS TO THE L12P FAMILY OF RIBOSOMAL PROTEINS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; M31339; AAA35336.1; -;
 CC EMBL; AL022070; CAA17793.1; -;
 CC PIR; C34715; R6BVP3.
 CC InterPro; IPR001813; 60s_ribosomal.
 CC Pfam; PF00428; 60S_Ribosomal; 1.
 CC Ribosomal protein; Phosphorylation; Multigene family.
 CC KW
 CC SEQUENCE 110 AA; 11171 MW; 1734AC9779F5A891 CRC64;
 CC
 CC Query Match 3.7%; Score 8; DB 1; Length 110;
 CC Best Local Similarity 100.0%; Pred. No. 1.9;
 CC Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 CC
 CC QY 119 SAAAPAA 126
 CC Db 64 SAAAPAA 71
 CC
 CC RESULT 7
 CC RL7_CHLPN STANDARD; PRT; 128 AA.
 CC ID Q929A1; Q9JQ70;
 CC DT 30-MAY-2000 (Rel. 39, Created)
 CC DT 10-MAY-2000 (Rel. 39, Last sequence update)
 CC DT 16-OCT-2001 (Rel. 40, Last annotation update)
 CC DE 50S ribosomal protein L7/L12.
 CC GN RPL7 OR RL7 OR CPN0080 OR CP0695.
 CC OS Chlamydia pneumoniae (Chlamydia pneumoniae).
 CC OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
 CC OX NCBI_TaxID=83558;
 CC [1]
 CC RN
 CC RP SEQUENCE FROM N.A.
 CC RC STRAIN=CWL029;
 CC RX MEDLINE=99206606; PubMed=10192388;
 CC RA Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
 CC Olinger L., Grimwood J., Davis R.W., Stephens R.S.;
 CC "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";
 CC Nat. Genet. 21:385-389(1999).
 CC [2]
 CC RN
 CC RP SEQUENCE FROM N.A.
 CC RC STRAIN=AR39;
 CC RX MEDLINE=20150255; PubMed=10684935;
 CC RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
 CC White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,
 CC Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
 CC Gwin M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
 CC Eisen J., Fraser C.M.;
 CC "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
 CC pneumoniae AR39.";
 CC Nucleic Acids Res. 28:1397-1406(2000).
 CC [3]
 CC RN
 CC RP SEQUENCE FROM N.A.
 CC RC STRAIN=J138;
 CC RX MEDLINE=20330349; PubMed=10871362;
 CC RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
 CC Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
 CC "Comparison of whole genome sequences of Chlamydia pneumoniae J138

RT from Japan and CWL029 from USA.";
 RL Nucleic Acids Res. 28:2311-2314(2000).
 CC -!- FUNCTION: SEEMS TO BE THE BINDING SITE FOR SEVERAL OF THE FACTORS
 CC INVOLVED IN PROTEIN SYNTHESIS AND APPEARS TO BE ESSENTIAL FOR
 CC ACCURATE TRANSLATION (BY SIMILARITY).
 CC -!- SIMILARITY: BELONGS TO THE L12P FAMILY OF RIBOSOMAL PROTEINS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AE001593; AAD18233.1; -;
 CC EMBL; AE002228; AAF38503.1; -;
 CC EMBL; AF002545; BAA98290.1; -;
 CC HSSP; P02392; ICTF.
 CC PHCI-2DPAGE; Q929A1; -;
 CC TIGR; CP0695; -;
 CC InterPro; IPR000206; Ribosomal_L12.
 CC Pfam; PF00542; Ribosomal_L12; 1.
 CC ProDom; PD001326; Ribosomal_L12; 1.
 CC TIGRFAMS; TIGR00855; L12; 1.
 CC Ribosomal protein; Complete proteome.
 CC KW
 CC INIT_MET 0 BY SIMILARITY.
 CC SEQUENCE 128 AA; 13461 MW; 4E2F17A85B057CC CRC64;
 CC
 CC Query Match 3.7%; Score 8; DB 1; Length 128;
 CC Best Local Similarity 100.0%; Pred. No. 2.2;
 CC Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 CC
 CC QY 171 EAPVAAEP 178
 CC Db 49 EAPVAAEP 56
 CC
 CC RESULT 8
 CC CUP5_GALME STANDARD; PRT; 353 AA.
 CC ID CUP5_GALME
 CC AC Q24998;
 CC DT 15-DEC-1998 (Rel. 37, Created)
 CC DT 15-DEC-1998 (Rel. 37, Last sequence update)
 CC DT 15-JUL-1999 (Rel. 38, Last annotation update)
 CC DE Pupal cuticle protein PCP52 precursor (GMPCP52).
 CC GN PCP52.
 CC OS Galleria mellonella (Wax moth).
 CC OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
 CC Insecta; Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata;
 CC Ditrysia; Pyraloidea; Pyralidae; Galleriinae; Galleria.
 CC OX NCBI_TaxID=7137;
 CC [1]
 CC RN
 CC RP SEQUENCE FROM N.A.
 CC RC TISSUE=Epidermis;
 CC RX MEDLINE=95291282; PubMed=7773255;
 CC RA Kollberg U., Obermaier B., Hirsch H., Kelber G., Wolbert P.;
 CC "Expression cloning and characterization of a pupal cuticle protein
 CC cDNA of Galleria mellonella L.";
 CC Insect Biochem. Mol. Biol. 25:355-363(1995).
 CC RL
 CC CC -!- FUNCTION: COMPONENT OF THE CUTICLE OF THE PUPA OF GALLERIA
 CC MELLONELLA.
 CC -!- DEVELOPMENTAL STAGE: MAXIMAL EXPRESSION DURING THE FIRST DAY AFTER
 CC PUPAL ECDYSIS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC

```
DR EMBL; X77514; CAA54650.1; -. Signal.  
KW Structural protein; Cuticle; POTENTIAL.  
FT SIGNAL 1 15  
FT CHAIN 16 353 PUPAL CUTICLE PROTEIN PCP52.  
FT DOMAIN 235 243 POLY-ALA.  
FT SEQUENCE 353 AA; 35818 MW; C3A35B04EC049172 CRC64;  
  
Query Match 3.7%; Score 8; DB 1; Length 353;  
Best Local Similarity 100.0%; Pred. No. 5.2;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 119 SAAAPAA 126  
| | | | | | | |  
Db 236 SAAAPAA 243  
  
RESULT 9  
P53_MESAU  
ID P53_MESAU STANDARD; PRT; 396 AA.  
AC Q00366; P97276;  
DT 01-DEC-1992 (Rel. 24, Created)  
DT 01-DEC-1992 (Rel. 24, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Cellular tumor antigen p53 (Tumor suppressor p53).  
GN TP53.  
OS Mesocricetus auratus (Golden hamster).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;  
OC Mesocricetus.  
OX NCBI_TaxID=10036;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Syrian; TISSUE=Kidney;  
RX MEDLINE=92210007; PubMed=1555773;  
RA Legros Y., McIntyre P., Soussi T.;  
RT "The cDNA cloning and immunological characterization of hamster p53";  
RL Gene 112:247-250(1992).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Hou E.W., Wiseman R.;  
RL Submitted (Apr-1994) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: Acts as a tumor suppressor in many tumor types; induces  
CC growth arrest or apoptosis depending on the physiological  
CC circumstances and cell type. Involved in cell cycle regulation as  
CC a trans-activator that acts to negatively regulate cell division  
CC by controlling a set of genes required for this process. One of  
CC the activated genes is an inhibitor of cyclin-dependent kinases.  
CC Apoptosis induction seems to be mediated either by stimulation of  
CC BAX and FAS antigen expression, or by repression of Bcl-2  
CC expression.  
CC -!- SUBUNIT: Binds DNA as a homotetramer (By similarity).  
CC -!- SUBCELLULAR LOCATION: Nuclear.  
CC -!- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY  
CC OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED  
CC IN MANY TYPES OF CANCER.  
CC -!- SIMILARITY: BELONGS TO THE P53 FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC -----  
CC EMBL; M75144; AAA37085.1; -.  
DR EMBL; U07182; AAB41344.1; -.  
DR PIR; JH0633; JH0633.  
DR HSSP; P04637; 1TUP.  
DR InterPro; IPR002117; P53.  
DR Pfam; PF00870; P53; 1.  
DR PRINTS; PR00386; P53SUPPRESSOR.  
DR ProDom; PD002681; P53; 1.  
  
DR PROSITE; PS00348; P53; 1.  
KW Anti-oncogene; DNA-binding; Transcription regulation; Activator;  
FT DOMAIN 1 45  
FT CHAIN 16 353 PUPAL CUTICLE PROTEIN PCP52.  
FT DOMAIN 235 243 POLY-ALA.  
FT SEQUENCE 353 AA; 35818 MW; C3A35B04EC049172 CRC64;  
  
Query Match 3.7%; Score 8; DB 1; Length 396;  
Best Local Similarity 100.0%; Pred. No. 5.7;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 120 AAAAPAAP 127  
| | | | | | | |  
Db 66 AAAAPAAP 73  
  
RESULT 10  
ATP2_CHLRE  
ID ATP2_CHLRE STANDARD; PRT; 574 AA.  
AC P38482;  
DT 01-OCT-1994 (Rel. 30, Created)  
DT 01-OCT-1994 (Rel. 30, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE ATP synthase beta chain, mitochondrial precursor (EC 3.6.3.14).  
GN ATP2.  
OS Chlamydomonas reinhardtii.  
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;  
OC Chlamydomonadaceae; Chlamydomonas.  
OX NCBI_TaxID=3055;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=92353386; PubMed=1386535;  
RA Franzen L.-G., Falk G.;  
RT "Nucleotide sequence of cDNA clones encoding the beta subunit of  
RT mitochondrial ATP synthase from the green alga Chlamydomonas  
RT reinhardtii: the precursor protein encoded by the cDNA contains both  
RT an N-terminal presequence and a C-terminal extension.";  
RL Plant Mol. Biol. 19:771-780(1992).  
RN [2]  
RP STRUCTURE BY NMR OF 1-26.  
RX MEDLINE=96326639; PubMed=8706917;  
RA Lancelin J.-M., Gans P., Bouchayer E., Bally I., Arian G.J.,  
RA Jacquot J.-P.;  
RT "NMR structures of a mitochondrial transit peptide from the green  
RT alga Chlamydomonas reinhardtii.";  
RL FEBS Lett. 391:203-208(1996).  
CC -!- FUNCTION: PRODUCES ATP FROM ADP IN THE PRESENCE OF A PROTON  
CC GRADIENT ACROSS THE MEMBRANE. THE BETA CHAIN IS THE CATALYTIC  
CC SUBUNIT.  
CC -!- CATALYTIC ACTIVITY: ATP + H(2)O + H(+)(In) = ADP + phosphate +  
CC H(+)(Out).  
CC -!- SUBUNIT: P-TYPE ATPASES HAVE 2 COMPONENTS, CF(1) - THE CATALYTIC  
CC CORE - AND CF(0) - THE MEMBRANE PROTON CHANNEL. CF(1) HAS FIVE  
CC SUBUNITS: ALPHA(3), BETA(3), GAMMA(1), DELTA(1), EPSILON(1). CF(0)  
CC HAS THREE MAIN SUBUNITS: A, B AND C.  
CC -!- SUBCELLULAR LOCATION: Mitochondrial.  
CC -!- SIMILARITY: BELONGS TO THE ATPASE ALPHA/BETA CHAINS FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC -----  
CC EMBL; X61624; CAA43808.1; -.  
DR PIR; S23530; S23530.
```

DR HSPD; P00829; 1BWF.
DR InterPro; IPR000793; ATPase_a/bC.
DR InterPro; IPR004100; ATPase_a/bN.
DR InterPro; IPR000194; ATPase_a/bCentre.
DR Pfam; PF00006; ATP-synt_ab; 1.
DR Pfam; PF00306; ATP-synt_ab.C; 1.
DR Pfam; PF02874; ATP-synt_ab.N; 1.
DR TIGRFAMs; TIGR01039; atcpd; 1.
DR PROSITE; PS00152; ATPASE_ALPHA_BETA; 1.
KW ATP synthesis; CR(1); Hydrogen ion transport;
KW Hydrolase; ATP-binding; Mitochondrion; Transit peptide.
FT TRANSIT 1 26 MITOCHONDRION (POTENTIAL).
FT CHAIN 27 574 ATP SYNTHASE BETA CHAIN.
FT NP_BIND 183 190 ATP (BY SIMILARITY).
SQ SEQUENCE 574 AA; 61821 MW; 22B36C6D18FBCFE CRC64;
Query Match 3.7%; Score 8; DB 1; Length 574;
Best Local Similarity 100.0%; Pred. No. 7.8;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 130 AAEPAAAA 137
|||||
DB 26 AAEPAAAA 33
RESULT 11
MSPA_TREMA STANDARD; PRT; 575 AA.
ID MSPA_TREMA
AC Q92413;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Major outer membrane protein mspA precursor (Major sheath protein).
DE MSPA.
GN Treponema maltophilum.
OS Bacteria; Spirochaetales; Spirochaetaceae; Treponema.
OX NCBI_TaxID=51160;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 20-34.
RC STRAIN=ATCC 51939 / BR;
RX MEDLINE=99121045; PubMed=9922270;
RA Heuner K., Choi B.-K., Schade R., Moter A., Otto A., Goebel U.B.;
RT "Cloning and characterization of a gene (mspA) encoding the major
sheath protein of Treponema maltophilum ATCC 51939(T).";
RJ J. Bacteriol. 181:1025-1029(1999).
RL [2]
RN SUBCELLULAR LOCALIZATION.
RP PubMed=11313133;
RX Heuner K., Meltzer U., Choi B.-K., Goebel U.B.;
RT "Outer sheath associated proteins of the oral spirochete Treponema
maltophilum.";
RJ FEMS Microbiol. Lett. 197:187-193(2001).
RL CC -1- FUNCTION: Major component of the outer membrane sheath.
CC -1- SUBCELLULAR LOCATION: Outer membrane-associated.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
at the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
or send an email to license@isb-sib.ch).
CC
CC EMBL; Y17800; CAA76862.1; -.
DR Outer membrane; Signal.
KW SIGNAL 1 19
FT CHAIN 20 575 MAJOR OUTER MEMBRANE PROTEIN MSPA.
SQ SEQUENCE 575 AA; 62250 MW; C666B188F5DE16A0 CRC64;
Query Match 3.7%; Score 8; DB 1; Length 575;
Best Local Similarity 100.0%; Pred. No. 7.9;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 140 TATPVAEE 147
|||||
DB 552 TATPVAEE 559
RESULT 13
MSPA_TREMA STANDARD; PRT; 576 AA.
ID MSPA_TREMA
AC Q18473;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Ecdysone receptor (Ecdysteroid receptor) (20-hydroxy-ecdysone
receptor) (20E receptor) (HVECR).
DE ECR OR NR1H1.
GN Heliothis virescens (Noctuid moth) (Owllet moth).
OS Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata;
OC Diptera; Noctuoidea; Noctuidae; Heliothinae; Heliothis.
OX NCBI_TaxID=7102;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BR-;
RX MEDLINE=99457718; PubMed=10528411;
RA Martinez A., Scanlon D., Gross B., Perara S.C., Palli S.R.,
RA Greenland A.J., Windass J., Pongs O., Broad P., Jepson I.;
RT "Transcriptional activation of the cloned Heliothis virescens
(Lepidoptera) ecdysone receptor (HVECR) by muristeroneA.";
RJ Insect Biochem. Mol. Biol. 29:915-930(1999).
RL CC -1- FUNCTION: RECEPTOR FOR ECDYSONE. BINDS TO ECDYSONE RESPONSE
ELEMENTS (ECRES) (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTOR FAMILY.
CC NR1 SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
at the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
or send an email to license@isb-sib.ch).
CC
CC EMBL; Y09009; CAA70212.1; -.
DR HSPD; P20393; lAGY.
DR InterPro; IPR000536; Hormone_rec_lig.
DR InterPro; IPR001628; ZnF_C4steroid.
DR Pfam; PF00104; hormone_rec; 1.
DR Pfam; PF00105; zf-C4; 1.
DR PRINTS; PR00047; STROIDFINGER.
DR PRODOM; PD000035; ZnF_C4steroid; 1.
DR SMART; SM00430; HOL1; 1.
DR SMART; PS00399; ZnF_C4; 1.
DR PROSITE; PS0031; NUCLEAR_RECEPTOR; 1.
DR Receptor; Transcription regulation; DNA-binding; Nuclear protein;
KW Zinc-finger. 1 162 MODULATING (POTENTIAL).
FT DOMAIN 163 228 NUCLEAR RECEPTOR-TYPE.
FT ZN_FING 163 183 C4-TYPE.
FT ZN_FING 199 223 C4-TYPE.
FT DOMAIN 326 545 HORMONE-BINDING (POTENTIAL).
SQ SEQUENCE 576 AA; 64638 MW; D13EF77BF263A8 CRC64;
Query Match 3.7%; Score 8; DB 1; Length 576;
Best Local Similarity 100.0%; Pred. No. 7.9;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 140 TATPVAEE 147
|||||
DB 552 TATPVAEE 559
RESULT 13

```

KELC DROME
ID KELC DROME STANDARD; PRT; 1477 AA.
AC Q04652; Q04653; Q9VJA2;
DT 01-FEB-1994 (Rel. 28, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Ring canal kelch protein [Contains: Kelch short protein].
GN KEL OR CG7210.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RX MEDLINE=93201592; PubMed=8453663;
RA Xue F., Cooley L.;
RT "Kelch encodes a component of intercellular bridges in Drosophila egg
chambers.";
RL Cell 72:681-693(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkely;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Aqbayani A., An H.-J., Andrews-pfannkuch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA De Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K.J., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fessler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong E., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harstin D.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Houlton D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turnek R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasmann D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
RN [3]
RP CHARACTERIZATION.
RC TISSUE=Embryo;
RX MEDLINE=97236487; PubMed=9118811;
RA Robinson D.N., Cooley L.;
RT "Examination of the function of two kelch proteins generated by stop
codon suppression.";
RL Development 124:1405-1417(1997).
CC -!- FUNCTION: COMPONENT OF RING CANALS THAT REGULATES THE FLOW OF
CYTOPLASM BETWEEN CELLS. MAY BE INVOLVED IN THE REGULATION OF

```

```

CC CC CYTOPLASM FLOW FROM NURSE CELLS TO THE OOCYTE DURING OOGENESIS.
CC CC BINDS ACTIN.
CC CC -!- SUBCELLULAR LOCATION: INNER SURFACE OF CYTOPLASMIC BRIDGES OR RING
CC CC CANALS PRESENT IN EGG CHAMBERS. SUBCORTICALLY IN IMAGINAL DISK
CC CC EPITHELIA.
CC CC -!- TISSUE SPECIFICITY: BOTH PROTEINS ARE EXPRESSED IN OVARIES, MALE
CC CC TESTIS, OVARECTOMIZED FEMALES, CUTICLE, SALIVARY GLAND AND
CC CC IMAGINAL DISKS. KELCH ORF1 IS THE PREDOMINANT PROTEIN AND IS ALSO
CC CC EXPRESSED IN FAT BODIES. ON ENTRY INTO METAMORPHOSIS LEVELS OF
CC CC FULL LENGTH PROTEIN INCREASE IN TESTIS AND IMAGINAL DISKS.
CC CC -!- DEVELOPMENTAL STAGE: LARVAE, PUPAE AND ADULTS.
CC CC -!- SIMILARITY: CONTAINS 1 BTB/POZ DOMAIN.
CC CC -!- SIMILARITY: CONTAINS 6 KELCH REPEATS.
CC CC -!- CAUTION: REF.3 BELIEVES RESIDUE 690 IS A SELENOCYSTEINE.
CC CC -!- CAUTION: Ref.2 sequence differs from that shown due to erroneous
CC CC gene model prediction.
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
CC CC EMBL; L08483; AAA53471.1; -
CC CC EMBL; L08483; AAA53472.2; -
CC CC EMBL; AE003657; AAF53651.1; ALT_SEQ.
CC CC HSSP; Q05316; 1CS3.
CC CC FlyBase; FBgn0001301; Kel.
CC CC InterPro; IPR000210; BTB_POZ.
CC CC InterPro; IPR001798; Kelch.
CC CC Pfam; PF00651; BTB; 1.
CC CC Pfam; PF01344; Kelch; 6.
CC CC PRINTS; PF03501; KELCHREPEAT.
CC CC SMART; SM00225; BTB; 1.
CC CC PROSITE; PS00097; BTB; 1.
CC CC Cytoskeleton; Actin-binding; Selenium; Selenocysteine; Repeat.
CC CC CHAIN 1 1477
CC CC FT CHAIN 1 689
CC CC FT CHAIN 1 223
CC CC FT CHAIN 1 449
CC CC FT REPEAT 404 496
CC CC FT REPEAT 450 543
CC CC FT REPEAT 498 543
CC CC FT REPEAT 545 592
CC CC FT REPEAT 594 639
CC CC FT REPEAT 641 687
CC CC FT REPEAT 641 687
CC CC FT DOMAIN 18 28
CC CC FT DOMAIN 29 87
CC CC FT DOMAIN 29 87
CC CC FT DOMAIN 29 86
CC CC FT DOMAIN 78 83
CC CC FT SE_CYS 690 690
CC CC FT CONFLICT 493 493 V -> A (IN REF. 1).
CC CC FT CONFLICT 596 596 A -> R (IN REF. 1).
CC CC FT CONFLICT 824 824 P -> L (IN REF. 1).
CC CC FT CONFLICT 858 858 G -> D (IN REF. 1).
CC CC FT CONFLICT 1083 1083 A -> R (IN REF. 1).
CC CC FT CONFLICT 1086 1086 A -> G (IN REF. 1).
CC CC SQ SEQUENCE 1477 AA; 160086 MW; 4851EEAE9D9DBA47 CRC64;
CC CC
CC CC Query Match 3.7%; Score 8; DB 1; Length 1477;
CC CC Best Local Similarity 100.0%; Pred. No. 17;
CC CC Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CC CC
CC CC QY 120 AAAAPAAP 127
CC CC Db 740 AAAAPAAP 747
CC CC
CC CC RESULT 14
CC CC NTC2_MOUSE
CC CC ID NTC2_MOUSE STANDARD; PRT: 2470 AA.
CC CC AC Q35516; Q60941; Q06008;

```



```

FT DOMAIN 26 63 EGF-LIKE 1.
FT DOMAIN 64 102 EGF-LIKE 2.
FT DOMAIN 105 143 EGF-LIKE 3.
FT DOMAIN 144 180 EGF-LIKE 4.
FT DOMAIN 182 219 EGF-LIKE 5.
FT DOMAIN 221 256 EGF-LIKE 6.
FT DOMAIN 258 294 EGF-LIKE 7.
FT DOMAIN 296 334 EGF-LIKE 8.
FT DOMAIN 336 372 EGF-LIKE 9.
FT DOMAIN 373 411 EGF-LIKE 10.
FT DOMAIN 413 452 EGF-LIKE 11.
FT DOMAIN 454 490 EGF-LIKE 12.
FT DOMAIN 492 528 EGF-LIKE 13.
FT DOMAIN 530 566 EGF-LIKE 14.
FT DOMAIN 568 603 EGF-LIKE 15.
FT DOMAIN 605 641 EGF-LIKE 16.
FT DOMAIN 643 678 EGF-LIKE 17.
FT DOMAIN 680 716 EGF-LIKE 18.
FT DOMAIN 718 753 EGF-LIKE 19.
FT DOMAIN 755 791 EGF-LIKE 20.
FT DOMAIN 793 829 EGF-LIKE 21.
FT DOMAIN 831 869 EGF-LIKE 22.
FT DOMAIN 871 907 EGF-LIKE 23.
FT DOMAIN 909 945 EGF-LIKE 24.
FT DOMAIN 947 983 EGF-LIKE 25.
FT DOMAIN 1023 1059 EGF-LIKE 26.
FT DOMAIN 1061 1097 EGF-LIKE 27.
FT DOMAIN 1099 1145 EGF-LIKE 28.
FT DOMAIN 1147 1183 EGF-LIKE 29.
FT DOMAIN 1185 1221 EGF-LIKE 30.
FT DOMAIN 1223 1260 EGF-LIKE 31.
FT DOMAIN 1262 1300 EGF-LIKE 32.
FT DOMAIN 1302 1345 EGF-LIKE 33.
FT DOMAIN 1372 1410 EGF-LIKE 34.
FT REPEAT 1418 1454 LIN/NOTCH 1.
FT REPEAT 1501 1533 ANK 1.
FT REPEAT 1825 1869 ANK 2.
FT REPEAT 1874 1903 ANK 3.
FT REPEAT 1907 1937 ANK 4.
FT REPEAT 1941 1970 ANK 5.
FT REPEAT 1974 2003 ANK 6.
FT REPEAT 2007 2036 POLY-ALA.
FT DOMAIN 1645 1648 POLY-LEU.
FT DOMAIN 1992 1995 POLY-ALA.
FT DOMAIN 2183 2189 POLY-SER.
FT DOMAIN 2425 2428 POLY-GLY.
FT DOMAIN 2445 2450 BY SIMILARITY.
FT DISULFID 28 41 BY SIMILARITY.
FT DISULFID 35 51 BY SIMILARITY.
FT DISULFID 53 62 BY SIMILARITY.
FT DISULFID 68 79 BY SIMILARITY.

Query Match 3.78; Score 8; DB 1; Length 2470;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 120 AAAAPAAP 127
DB 2183 AAAAPAAP 2190

RESULT 15
ANPA_PSEAM STANDARD; PRT; 82 AA.
AC P04002;
DT 23-OCT-1986 (Rel. 02, Created)
DT 23-OCT-1986 (Rel. 02, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Antifreeze protein A/B precursor.
OS Pseudopleuronectes americanus (winter flounder) (Pleuronectes americanus).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Pleuronectiformes;
OC Pleuronectoidae; Pleuronectidae; Pseudopleuronectes.
OX NCBI_taxid=8265;
RN [1]
RP SEQUENCE FROM N.A. (PROTEIN A).
RX MEDLINE=82197490; PubMed=6952188;
RA Davies P.L., Roach A.H., Hew C.-L.;
RT "DNA sequence coding for an antifreeze protein precursor from winter flounder.";
RL Proc. Natl. Acad. Sci. U.S.A. 79:335-339(1982).
RN [2]
RP SEQUENCE FROM N.A. (PROTEIN A).
RX MEDLINE=88259236; PubMed=3133486;
RA Scott G.K., Davies P.L., Kao M.H., Fletcher G.L.;
RT "Differential amplification of antifreeze protein genes in the pleuronectinae.";
RL J. Mol. Evol. 27:29-35(1988).
RN [3]
RP SEQUENCE FROM N.A. (PROTEIN B).
RX MEDLINE=84264559; PubMed=6086629;
RA Davies P.L., Hough C., Scott G.K., Ng N., White B.N., Hew C.-L.;
RT "Antifreeze protein genes of the winter flounder.";
RL J. Biol. Chem. 259:9241-9247(1984).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=92209995; PubMed=1555765;
RA Davies P.L.;
RT "Conservation of antifreeze protein-encoding genes in tandem repeats.";
RL Gene 112:163-170(1992).
RN [5]
RP 3D-STRUCTURE MODELING OF 45-81.
RX MEDLINE=92148833; PubMed=1738160;
RA Chou K.-C.;
RT "Energy-optimized structure of antifreeze protein and its binding mechanism.";
RL J. Mol. Biol. 223:509-517(1992).
RN [6]
RP X-RAY CRYSTALLOGRAPHY (1.5 ANGSTROMS) OF 45-81.
RX MEDLINE=95281060; PubMed=7760940;
RA Siccheri F., Yang D.S.C.;
RT "Ice-binding structure and mechanism of an antifreeze protein from winter flounder.";
RL Nature 375:427-431(1995).
CC -!- FUNCTION: ANTIFREEZE PROTEINS LOWER THE BLOOD FREEZING POINT.
CC -!- MISCELLANEOUS: THE SEQUENCE SHOWN IS THAT OF PROTEIN A.
CC -!- SIMILARITY: BELONGS TO THE TYPE-1 AFP FAMILY. TYPE 1 AFP ARE ALANINE-RICH, AMPHIPHILIC AND ALPHA-HELICAL.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L00138; AAB59964.1; -
CC EMBL; L29178; AAB59964.1; JOINED.
CC EMBL; M62414; AAA49469.1; -
CC EMBL; X07506; CAA30389.1; -
CC EMBL; M62416; AAA49471.1; -
CC EMBL; M62417; AAA49472.1; -
CC PIR; A03194; FDFLAW.
CC PIR; A05161; A05161.
CC PIR; S02326; S02326.
CC PIR; JS0704; JS0704.
CC PDB; 1ATF; 15-OCT-94.
CC PDB; 1WFA; 03-JUN-95.
CC PDB; 1WFB; 03-JUN-95.
CC InterPro; IPR00104; Antifreeze_1.
CC PRINTS; PR00308; ANTIFREEZE1.

```

```
KW Antifreeze protein; Repeat; Multigene family; Signal; 3D-structure.
FT SIGNAL 1 21
FT PROPEP 22 44
FT CHAIN 45 82
FT VARIAT 36 36
FT VARIAT 70 70
FT CONFLICT 24 24
FT HELIX 46 80
SQ SEQUENCE 82 AA; 7711 MW; C2AE7B74C0D46CC1 CRC64;

Query Match 3.2%; Score 7; DB 1; Length 82;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 120 AAAPAA 126
Db 35 AAAPAA 41
```

Search completed: February 12, 2003, 13:58:20
Job time : 18 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2003, 13:53:55 ; Search time 31 Seconds
(without alignments)
1455.623 Million cell updates/sec

Title: US-09-923-236-2

Perfect score: 219

Sequence: 1 MKLLWACIVCVAFARRR.....TAKPAPEHPSPSLEQANQ 219

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 671580 seqs, 206047115 residues

Word size : 0

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

1: SP_TREMBL_21.*

2: sp_archaea.*

3: sp_bacteria.*

4: sp_fungi.*

5: sp_human.*

6: sp_invertebrate.*

7: sp_mammal.*

8: sp_mhc.*

9: sp_organelle.*

10: sp_phase.*

11: sp_plant.*

12: sp_rhodent.*

13: sp_virus.*

14: sp_vertebrate.*

15: sp_unclassified.*

16: sp_virus.*

17: sp_bacteriap.*

18: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	9	4.1	75	Q9VBL6	Q9vbl6 drosophila
2	9	4.1	344	Q9D5J3	Q9d5j3 mus musculus
3	9	4.1	728	Q9VY77	Q9vy77 drosophila
4	9	4.1	841	Q9NEA5	Q9nea5 leishmania
5	8	3.7	139	Q8Y2D2	Q8y2d2 raietonia s
6	8	3.7	141	Q8Y2R3	Q8y2r3 raietonia s
7	8	3.7	145	Q96516	Q96516 arabidopsis
8	8	3.7	145	Q9FFC0	Q9ffc0 arabidopsis
9	8	3.7	193	Q9C7F7	Q9c7f7 arabidopsis
10	8	3.7	209	Q9D0U8	Q9d0u8 human immun
11	8	3.7	228	Q9ADA2	Q9ada2 caulobacter
12	8	3.7	231	Q9SND2	Q9snd2 myxococcus
13	8	3.7	259	Q91UQ4	Q91uq4 rhizobium m
14	8	3.7	261	Q9CSG8	Q9csg8 mus musculus
15	8	3.7	262	Q8W2Q2	Q8w2q2 oryza sativ
16	8	3.7	265	Q9CSK8	Q9csk8 mus musculus

17	8	3.7	268	16	P95192	P95192 mycobacteri
18	8	3.7	272	10	Q8S2S9	Q8s2s9 thellungiel
19	8	3.7	276	16	Q92WH2	Q92wh2 rhizobium m
20	8	3.7	291	10	Q40478	Q40478 nicotiana t
21	8	3.7	302	5	Q9N9A5	Q9n9a5 leishmania s
22	8	3.7	325	2	Q929X8	Q929x8 frateuria s
23	8	3.7	336	16	Q9AB24	Q9ab24 caulobacter
24	8	3.7	362	11	Q9CX00	Q9cx00 mus musculu
25	8	3.7	381	2	Q86996	Q86996 acinetobact
26	8	3.7	384	16	Q9AD02	Q9ad02 streptomyce
27	8	3.7	428	10	Q94IQ8	Q94iq8 zea mays (m
28	8	3.7	447	8	Q9TN70	Q9tn70 tupistra al
29	8	3.7	448	8	Q33130	Q33130 streitizia
30	8	3.7	468	8	Q9GEQ0	Q9geq0 primula cor
31	8	3.7	493	2	Q48431	Q48431 klebsiella
32	8	3.7	493	5	Q9W4P4	Q9w4p4 drosophila
33	8	3.7	493	5	Q8SZR1	Q8szr1 drosophila
34	8	3.7	493	10	Q9XEN0	Q9xen0 oryza sativ
35	8	3.7	512	5	Q9VMI4	Q9vmi4 drosophila
36	8	3.7	550	16	Q9RY21	Q9ry21 deinococcus
37	8	3.7	551	3	Q42691	Q42691 alternaria
38	8	3.7	555	2	Q85737	Q85737 streptomyce
39	8	3.7	569	10	Q8W2Q4	Q8w2q4 oryza sativ
40	8	3.7	600	16	Q8XSG9	Q8xsg9 raietonia s
41	8	3.7	611	16	Q92TK5	Q92tk5 rhizobium m
42	8	3.7	679	2	Q93AS9	Q93as9 raietonia s
43	8	3.7	715	11	Q63803	Q63803 rattus norv
44	8	3.7	789	16	Q8XX15	Q8xx15 raietonia s
45	8	3.7	842	10	Q9LEE7	Q9lee7 zea mays (m

ALIGNMENTS

RESULT 1

Q9VBL6	PRELIMINARY;	PRT;	75 AA.
ID	Q9VBL6	PRELIMINARY;	PRT;
AC	Q9VBL6: Q24390;		
DC	01-MAY-2000 (TREMBLrel. 13, Created)		
DT	01-MAY-2000 (TREMBLrel. 13, Last sequence update)		
DE	01-JUN-2001 (TREMBLrel. 17, Last annotation update)		
DE	Accessory gland-specific peptide 57Da precursor (Male accessory gland		
DE	secretory protein 57Da).		
GN	MST57DA OR CG9074.		
OS	Drosophila melanogaster (Fruit fly).		
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;		
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;		
OC	Ephydroidea; Drosophilidae; Drosophila.		
OX	NCBI_TaxID=7227;		
RN	[1]		
RP	SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.		
RC	STRAIN=OREGON-R;		
RX	MEDLINE=95227188; PubMed=7711745;		
RA	Simmerl E., Schaefer M., Schaefer U.;		
RT	"Structure and regulation of a gene cluster for male accessory gland		
RT	transcripts in Drosophila melanogaster."		
RL	Insect Biochem. Mol. Biol. 25:127-137(1995).		
RL	[2]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=BERKELEY;		
RX	MEDLINE=20196006; PubMed=10731132;		
RA	Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,		
RA	Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,		
RA	George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,		
RA	Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,		
RA	Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,		
RA	Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,		
RA	Abril J.F., Abmayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,		
RA	Balieu R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,		
RA	Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,		
RA	Borkova D., Bouchan M.R., Bouck J., Brokstein P., Brotter P.,		
RA	Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,		
RA	Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,		

RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulo G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Patzoldo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.:
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
CC -!- FUNCTION: THIS PROTEIN IS TRANSFERRED FROM MALE TO FEMALE DURING
CC MATING AND MAY AFFECT EGG-LAYING AND BEHAVIOR AFTER MATING.
CC -!- SUBCELLULAR LOCATION: SECRETED.
CC -!- TISSUE SPECIFICITY: LUMEN FLUID OF MALE ACCESSORY GLANDS, BECOMES
CC SEMINAL FLUID.
DR EMBL: Z33647; CAA83925.1; -.
DR EMBL: AE003753; AAF56515.1; -.
DR FlyBase: FBgn0011668; Mst57Da.
KW Signal; Behavior.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 75 ACCESSORY GLAND-SPECIFIC PEPTIDE 57DA.
FT DOMAIN 39 64 ALA/PRO-RICH DOMAIN.
FT CONFLICT 39 46 MISSING (IN REF. 1).
FT CONFLICT 64 75 MISSING (IN REF. 1).
FT CONFLICT 75 75 BCD3D70817C98E14 CRC64;
SQ SEQUENCE 75 AA; 7106 MW; BCD3D70817C98E14 CRC64;

Query Match 4.1%; Score 9; DB 5; Length 75;
Best Local Similarity 100.0%; Pred. No. 0.65;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 120 AAAAPAAP 128
DB 45 AAAAPAAP 53

RESULT 2
ID Q9D5J3 PRELIMINARY; PRT; 344 AA.
AC Q9D5J3;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE 4930432K09Rik protein.
GN 4930432K09Rik.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=TESTIS;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,

RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Kamiya M., Lee N.H.,
RA Gystincich S., Hill D., Hofmann M., Hume D.A., Mazzarelli J., Monbaerts P.,
RA Lyons P., Marchionni L., Mashima J., Rodriguez I., Sakamoto N.,
RA Nordone P., Ring B., Ringwald M., Seva T., Shibata Y., Storch K.-F.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Whittaker C., Wilming L.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Yawshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection."
RL Nature 409:685-690(2001).
DR EMBL: AK015291; BAB29782.1; -.
DR MGD: MGI:1921029; 4930432K09Rik.
DR InterPro: IPR002965; P-rich_extensn.
DR PRINTS: PRO1217; PRICHEXTENSN.
SQ SEQUENCE 344 AA; 36294 MW; D65EAD71CE5802AC CRC64;

Query Match 4.1%; Score 9; DB 11; Length 344;
Best Local Similarity 100.0%; Pred. No. 2.5;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLLWACI 9
DB 1 MKLLWACI 9

RESULT 3
Q9VY77
ID Q9VY77 PRELIMINARY; PRT; 728 AA.
AC Q9VY77;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE CG11063 protein.
GN CG11063.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulo G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Muzny D.M., Nelson D.L.,

```

RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
CC -!- SIMILARITY: CONTAINS 3 LIM DOMAINS. THE LIM DOMAIN BINDS 2 ZINC
IONS.
DR EMBL; AE003493; AAF48328.1; -.
DR HSSP; P04006; 1IML.
DR FlyBase; FBgn0030530; CG11063.
DR InterPro; IPR001781; LIM.
DR Pfam; PF00412; LIM; 3.
DR ProDom; PD000094; LIM; 3.
DR SMART; SM00132; LIM; 3.
DR PROSITE; PS00478; LIM_DOMAIN.1; 2.
DR PROSITE; PS0023; LIM_DOMAIN.2; 3.
KW LIM domain; Metal-binding; Zinc.
SQ SEQUENCE 728 AA; 77961 MW; 86F9CAFF840E4541 CRC64;

Query Match 4.1%; Score 9; DB 5; Length 728;
Best Local Similarity 100.0%; Pred. No. 5;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 119 SAAAPAAP 127
Db 418 SAAAPAAP 426
|||||

RESULT 4
Q9NEA5 PRELIMINARY; PRT; 841 AA.
ID Q9NEA5;
AC Q9NEA5;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Possible inner dynein arm.
GN L5515.03.
OS Leishmania major.
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=5664;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-FRIEDLIN;
RA Masuy D., Purnelle B., Goffeau A., Ivens A.C., Quail M.,
RA Rajandream M.A., Barrell B.G.;
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-FRIEDLIN;
RX MEDLINE=98146435; PubMed=9477341;
RA Ivens A.C., Lewis S.M., Badgerzadeh A., Zhang L., Chan H.M.,
RA Smith D.F.;
RT "A physical map of the Leishmania major Friedlin genome.";
RL Genome Res. 8:135-145(1998).
CC -!- SIMILARITY: CONTAINS 4 WD REPEATS (TRP-ASP DOMAINS).
DR EMBL; AL161414; CAB77677.1; -.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 4.
DR SMART; SM00320; WD40; 4.
DR PROSITE; PS00678; WD_REPEATS.1; UNKNOWN.2.
DR PROSITE; PS0294; WD_REPEATS_REGION; 1.
KW Repeat; WD repeat.
SQ SEQUENCE 841 AA; 93125 MW; B41B72F1BE821226 CRC64;

Query Match 4.1%; Score 9; DB 5; Length 841;

```

```

Best Local Similarity 100.0%; Pred. No. 5.7;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 119 SAAAPAAP 127
Db 40 SAAAPAAP 48
|||||

RESULT 5
Q8Y2D2 PRELIMINARY; PRT; 139 AA.
ID Q8Y2D2;
AC Q8Y2D2;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DE Hypothetical signal peptide protein RSC0404.
GN RSC0404 OR RS03372.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
OC Ralstonia.
OX NCBI_TaxID=305;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-GMI1000;
RX MEDLINE=21681879; PubMed=11823852;
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA Arlat M., Billault A., Bottier P., Camus J.C., Cattolico L.,
RA Chandler M., Choisme N., Claudel-Renard C., Cunnac S., Demange N.,
RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,
RA Sigquier P., Thebault P., Whalen M., Wincker P., Levy M.,
RA Weissbach J., Boucher C.A.;
RT "Genome sequence of the plant pathogen Ralstonia solanacearum.";
RL Nature 415:497-502(2002).
RL EMBL; AL646059; CAD13932.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 139 AA; 14153 MW; A31350B6FDG29039 CRC64;

Query Match 3.7%; Score 8; DB 16; Length 139;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 120 AAAAPAAP 127
Db 55 AAAAPAAP 62
|||||

RESULT 6
Q8Y2R3 PRELIMINARY; PRT; 141 AA.
ID Q8Y2R3;
AC Q8Y2R3;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical protein RSC0272.
GN RSC0272 OR RS03239.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
OC Ralstonia.
OX NCBI_TaxID=305;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-GMI1000;
RX MEDLINE=21681879; PubMed=11823852;
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA Arlat M., Billault A., Bottier P., Camus J.C., Cattolico L.,
RA Chandler M., Choisme N., Claudel-Renard C., Cunnac S., Demange N.,
RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,
RA Sigquier P., Thebault P., Whalen M., Wincker P., Levy M.,
RA Weissbach J., Boucher C.A.;
RT "Genome sequence of the plant pathogen Ralstonia solanacearum.";
RL Nature 415:497-502(2002).
RL EMBL; AL646058; CAD13800.1; -.
KW InterPro; IPR001763; Rhodanese-like.

```

DR	SMART; SM00450; RHOD; 1.
KW	Hypothetical protein; Complete proteome.
SQ	SEQUENCE 141 AA; 13454 MW; 878FE1840A8E8B4E CRC64;
Query Match	3.7%; Score 8; DB 16; Length 141;
Best Local Similarity	100.0%; Pred. No. 10;
Matches	8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	149 AAGAPVAA 156
Dd	126 AAGAPVAA 133
RESULT 7	
Q96516	PRELIMINARY; PRT; 145 AA.
ID	Q96516
AC	Q96516;
DT	01-FEB-1997 (TrEMBLrel. 02, Created)
DT	01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT	01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE	Histone H2B.
OS	Arabidopsis thaliana (Mouse-ear cress).
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC	eurosid II; Brassicales; Brassicaceae; Arabidopsids.
OX	NCBI_TaxID=3702;
FN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=ECOTYPE COLUMBIA; TISSUE=CELL SUSPENSION CULTURE;
RA	Phillips G.;
RL	Submitted (JUL-1996) to the EMBL/GenBank/DBDJ databases.
CC	-1- SUBUNIT: THE NUCLEOSOME IS AN OCTAMER CONTAINING TWO MOLECULES OF
CC	H2A, H2B, H3, AND H4; WHICH WRAP APPROXIMATELY 146 BP OF DNA (BY
CC	SIMILARITY).
CC	-1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC	-1- SIMILARITY: BELONGS TO THE HISTONE H2B FAMILY.
DR	EMBL; Y07745; CAA69025.1; -.
DR	InterPro; IPR004822; Histone_core.
DR	InterPro; IPR000558; Histone_H2B.
DR	Pfam; PF00125; histone; 1.
DR	PRINTS; PR00621; HISTONEH2B.
DR	PRODom; PD000497; Histone_H2B; 1.
DR	SMART; SM00427; H2B; 1.
DR	PROSITE; PS00357; HISTONE_H2B; 1.
KW	Chromosomal protein; DNA-binding; Nuclear protein; Nucleosome core.
SQ	SEQUENCE 145 AA; 15733 MW; 6C8E8B18390F4686 CRC64;
Query Match	3.7%; Score 8; DB 10; Length 145;
Best Local Similarity	100.08; Pred. No. 11;
Matches	8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	130 AAEPAAAA 137
Dd	19 AAEPAAAA 26
RESULT 8	
Q9FFCO	PRELIMINARY; PRT; 145 AA.
ID	Q9FFCO
AC	Q9FFCO;
DT	01-WAR-2001 (TrEMBLrel. 16, Created)
DT	01-WAR-2001 (TrEMBLrel. 16, Last sequence update)
DT	01-WAR-2002 (TrEMBLrel. 20, Last annotation update)
DE	Histone H2B.
OS	Arabidopsis thaliana (Mouse-ear cress).
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC	eurosid II; Brassicales; Brassicaceae; Arabidopsids.
OX	NCBI_TaxID=3702;
FN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=COLUMBIA;
RX	MEDLINE=97471969; PubMed=9330910;

```

RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,
RA Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,
RA Ecker J., Theologis A., Davis R.W.,
RA Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC069471; AAG51485.1; -.
DR EMBL; AY092956; AAM12955.1; -.
DR InterPro; IPR003612; AAI.
DR Pfam; PF00234; tryp_alpha_amyl; 1.
DR SMART; SM00499; AAI_1.
DR SEQUENCE 193 AA; 19759 MW; D54B38B12FPE6610 CRC64;

Query Match 3.7%; Score 8; DB 10; Length 193;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 120 AAAAPAAP 127
DB 20 AAAAPAAP 27

RESULT 10
Q9DQ08 PRELIMINARY; PRT; 209 AA.
AC Q9DQ08;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Negative factor (F-protein) (27 kDa protein).
GN NEF.
OS Human immunodeficiency virus type 1.
OC Viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=01U09;
RX MEDLINE=21002575; PubMed=11118071;
RA Geffin R., Wolf D., Muller R., Hill M.D., Stellwag E., Freitag M.,
RA Sass G., Scott G.B., Baur A.S.;
RT "Functional and structural defects in HIV-1 nef genes derived from
RT pediatric long-term survivors.";
RL AIDS Res. Hum. Retroviruses 16:1855-1869(2000).
CC -!- SIMILARITY: NEF HAS GTPASE, GTP-BINDING AND AUTOPHOSPHORYLATING
CC ACTIVITIES, IT SEEMS TO DOWN-REGULATE THE CD4 (T4) ANTIGEN.
DR EMBL; AF252901; AAG34573.1; -.
DR HSP; P03406; 1EFN.
DR InterPro; IPR001558; HIV_Nef.
DR Pfam; PF00469; F-protein; 1.
DR ProDom; PD000031; HIV_Nef; 1.
DR AIDS; GTP-binding; Myristate.
DR NCBI_TaxID=34;
DR SEQUENCE 209 AA; 23656 MW; 6B858BAD98C79519 CRC64;

Query Match 3.7%; Score 8; DB 15; Length 209;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 131 AEPAAAP 138
DB 23 AEPAAAP 30

RESULT 11
Q9A4A2 PRELIMINARY; PRT; 228 AA.
AC Q9A4A2;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Cytochrome c, membrane-bound.
GN CC2935.
OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
OC Caulobacter.

```

```

OX NCBI_TaxID=155892;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 19089 / CBL15;
RX MEDLINE=21173698; PubMed=11259647;
RA Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA Klotterback J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA Ullery T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT "Complete genome sequence of Caulobacter crescentus.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
DR EMBL; AE005958; AAK24897.1; -.
DR HSP; P81459; 1I54.
DR TIGR; CC2935; -.
DR InterPro; IPR00104; Antifreeze1.
DR InterPro; IPR000345; CytC_heme_bind.
DR InterPro; IPR003088; Cyt_C1.
DR Pfam; PF00034; cytochrome_c; 1.
DR PRINTS; PR00308; ANTIFREEZE1.
DR PRINTS; PR00604; CYTCHRMCIAB.
DR ProDom; PD000375; Cyt_C1AB; 1.
DR PROSITE; PS00190; CYTOCHROME_C; UNKNOWN_1.
DR Complete proteome.
DR SEQUENCE 228 AA; 23023 MW; 5AE05C7DD5D05F82 CRC64;

Query Match 3.7%; Score 8; DB 16; Length 228;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 120 AAAAPAAP 127
DB 219 AAAAPAAP 226

RESULT 12
Q93ND2 PRELIMINARY; PRT; 231 AA.
AC Q93ND2;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE Hypothetical 24.5 kDa protein.
OS Myxococcus xanthus.
OC Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
OC Myxococcales; Cytopharyngae; Myxococcaceae; Myxococcus.
OX NCBI_TaxID=34;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DZFL;
RA Nariya H., Inouye S.;
RT "Identification of serine/threonine kinase associate proteins in M.
RT xanthus by yeast two-hybrid system.";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF377338; AAK64438.1; -.
DR Hypothetical protein.
DR SEQUENCE 231 AA; 24488 MW; 8A2AA9FF8D789536 CRC64;

Query Match 3.7%; Score 8; DB 2; Length 231;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 149 AAGAPVAA 156
DB 39 AAGAPVAA 46

RESULT 13
Q91UQ4 PRELIMINARY; PRT; 259 AA.
ID Q91UQ4

```

```
AC Q91U04;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE VirB1 protein.
GN VIRB1.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OG Plasmid pSB102.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RC TRANSPOSON-MERCURY RESISTANCE TRANSPOSON TN5718;
RA Schneider S., Keller M., Droegge M., Lanka E., Puehler A.,
RA Selbitschka W.;
RT "The genetic organization and evolution of the broad-host-range
RT mercury resistance plasmid pSB102 isolated from a microbial population
RT residing in the rhizosphere of alfalfa.";
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
RL EMBL; AJ304453; CAC79183.1; -.
DR InterPro: IPR001179; FKBP_PPIase.
DR Pfam: PF01464; SLT; 1.
DR PROSITE: PS00453; FKBP_PPIASE_1; UNKNOWN_1.
KW Plasmid.
SQ SEQUENCE 259 AA; 27380 MW; 09648108175E9A7B CRC64;

Query Match 3.7%; Score 8; DB 2; Length 259;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 152 APVAAEPA 159
DB 193 APVAAEPA 200
|||||||

RESULT 14
Q9CSG8 PRELIMINARY; PRT; 261 AA.
AC Q9CSG8;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE 1810019E15RIK protein (Fragment).
GN 1810019E15RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=EMBRYO;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gofobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
```

```
RL Nature 409:685-690(2001).
DR EMBL; AK012877; BAB28530.1; -.
DR MGI; 1914254; 1810019E15RIK.
DR InterPro: IPR003034; SAP.
DR Pfam; PF02037; SAP; 1.
DR SMART; SM00513; SAP; 1.
FT NON_TER 261
SQ SEQUENCE 261 AA; 29664 MW; D5CA7CDC443F4A9E CRC64;
```

```
Query Match 3.7%; Score 8; DB 11; Length 261;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 119 SAAAPAA 126
DB 2 SAAAPAA 9
|||||||
```

RESULT 15

```
Q8W2Q2 PRELIMINARY; PRT; 262 AA.
AC Q8W2Q2;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Putative polyprotein.
GN OSJNBA0028C16.20.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RA Wang R.A., Yu Y., Soderlund C., Chen M., Kim H.-R., Rambo T.,
RA Saski C., Henry D., Oates R., Simmons J.;
RT "Rice Genomic Sequence.";
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC098565; AAL69438.1; -.
KW Polyprotein.
SQ SEQUENCE 262 AA; 27461 MW; C7BA920CF2486429 CRC64;
```

```
Query Match 3.7%; Score 8; DB 10; Length 262;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 121 AAAPAAP 128
DB 148 AAAPAAP 155
|||||||
```

Search completed: February 12, 2003, 13:59:04
Job time : 33 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2003, 11:53:14 ; Search time 65 Seconds
(without alignments)
448.952 Million cell updates/sec

Title: US-09-923-236-2
Perfect score: 1178
Sequence: 1 MKLLWACIVCVAFKRRF.....TAKPAPEPHSPSLEQANQ 219

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002: *
1: /SID22/gcgdata/geneseq/geneseq-emb1/AA1980.DAT: *
2: /SID22/gcgdata/geneseq/geneseq-emb1/AA1981.DAT: *
3: /SID22/gcgdata/geneseq/geneseq-emb1/AA1982.DAT: *
4: /SID22/gcgdata/geneseq/geneseq-emb1/AA1983.DAT: *
5: /SID22/gcgdata/geneseq/geneseq-emb1/AA1984.DAT: *
6: /SID22/gcgdata/geneseq/geneseq-emb1/AA1985.DAT: *
7: /SID22/gcgdata/geneseq/geneseq-emb1/AA1986.DAT: *
8: /SID22/gcgdata/geneseq/geneseq-emb1/AA1987.DAT: *
9: /SID22/gcgdata/geneseq/geneseq-emb1/AA1988.DAT: *
10: /SID22/gcgdata/geneseq/geneseq-emb1/AA1989.DAT: *
11: /SID22/gcgdata/geneseq/geneseq-emb1/AA1990.DAT: *
12: /SID22/gcgdata/geneseq/geneseq-emb1/AA1991.DAT: *
13: /SID22/gcgdata/geneseq/geneseq-emb1/AA1992.DAT: *
14: /SID22/gcgdata/geneseq/geneseq-emb1/AA1993.DAT: *
15: /SID22/gcgdata/geneseq/geneseq-emb1/AA1994.DAT: *
16: /SID22/gcgdata/geneseq/geneseq-emb1/AA1995.DAT: *
17: /SID22/gcgdata/geneseq/geneseq-emb1/AA1996.DAT: *
18: /SID22/gcgdata/geneseq/geneseq-emb1/AA1997.DAT: *
19: /SID22/gcgdata/geneseq/geneseq-emb1/AA1998.DAT: *
20: /SID22/gcgdata/geneseq/geneseq-emb1/AA1999.DAT: *
21: /SID22/gcgdata/geneseq/geneseq-emb1/AA2000.DAT: *
22: /SID22/gcgdata/geneseq/geneseq-emb1/AA2001.DAT: *
23: /SID22/gcgdata/geneseq/geneseq-emb1/AA2002.DAT: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1178	100.0	219	22	Human transport pr
2	1178	100.0	219	23	Human zsig63 polyp
3	1172	99.5	219	20	Human secreted pro
4	1113.5	94.5	221	20	Amino acid sequenc
5	183.5	15.6	325	22	Novel human diagno
6	180	15.3	207	14	Sequence of a 10-a
7	179	15.2	207	17	T. cruzi Tcd antigen
8	179	15.2	207	20	Tcd protein sequen
9	179	15.2	207	20	Tcd antigen of Trp
10	175.5	14.9	718	22	Human polypeptide

11	175.5	14.9	903	23	ABP411277	Human ovarian anti
12	175.5	14.9	3177	22	AAW40303	Human polypeptide
13	174	14.8	566	22	ABB61040	Drosophila melanog
14	173	14.7	538	22	AAW82806	Human low density
15	170.5	14.5	149	22	AAU00451	Protein encoded by
16	170	14.4	267	19	AAW81726	M. tuberculosis im
17	170	14.4	267	19	AAW64359	Mycobacterium tube
18	170	14.4	267	20	AAW39156	M. tuberculosis an
19	170	14.4	267	20	AAW39013	M. tuberculosis re
20	170	14.4	428	23	AAW50963	Maize methyl Cpg b
21	170	14.4	842	22	ABB66631	Drosophila melanog
22	170	14.4	864	22	ABB71319	Drosophila melanog
23	170	14.4	3173	23	ABB90743	Human Tumour Endot
24	169.5	14.4	783	19	AAW37151	Mouse neural Mena+
25	169.5	14.4	787	19	AAW37152	Mouse neural Mena+
26	169.5	14.4	802	19	AAW37153	Mouse neural Mena+
27	169.5	14.4	802	22	AAU09139	Mammalian enabled
28	169	14.3	299	22	ABB63276	Drosophila melanog
29	168.5	14.3	550	22	AAW82807	Rabbit low density
30	168.5	14.3	706	22	ABB66693	Drosophila melanog
31	165.5	14.0	763	18	AAW31852	Mycobacterium tube
32	164	13.9	572	18	AAW31855	Mycobacterium tube
33	163.5	13.9	214	17	AAW86913	Cotton fiber speci
34	162.5	13.8	1064	22	ABB61785	Drosophila melanog
35	162	13.8	471	22	ABB66662	Drosophila melanog
36	159.5	13.5	203	23	ABG60116	Human D1THP polype
37	159.5	13.5	421	22	ABB71817	Drosophila melanog
38	159.5	13.5	537	23	AAU82746	Amino acid sequenc
39	156.5	13.3	681	23	ABW93650	Herbicidally activ
40	156	13.2	659	22	ABB65280	Drosophila melanog
41	156	13.2	3119	19	AAW72204	HSV-2 strain SB5 C
42	155	13.2	574	22	AAW52322	WASP homolog prote
43	155	13.2	574	22	AAW52322	Amino acid sequenc
44	155	13.2	675	21	AAW54052	An angiogenesis-as
45	155	13.2	675	21	AAW54053	A variant of an an

ALIGNMENTS

RESULT 1
AAB60109
ID AAB60109 standard; Protein; 219 AA.

AC AAB60109;

DT 28-MAR-2001 (first entry)

DE Human transport protein TPPT-29.

DE Human transport protein TPPT-29.
KW Human; transport protein; TPPT; transport disorder; metabolic disorder;
KW neurological disorder; cardiovascular disorder; reproductive disorder;
KW immune disorder; cancer.

OS Homo sapiens.

PN WO200078953-A2.

PD 28-DEC-2000.

PF 16-JUN-2000; 2000WO-US16668.

PR 17-JUN-1999; 99US-0139923.

PR 10-AUG-1999; 99US-0148177.

PR 18-AUG-1999; 99US-0149357.

PR 28-OCT-1999; 99US-0162287.

PA (INCY-) INCYTE GENOMICS INC.

PI Lal P, Yang J, Yue H, Hillman JL, Tang YT, Bandman O, Burford N;
PI Baughn MR, Azimzai Y, Lu DAM, Au-Young J, Patterson C;
PI WPI; 2001-041424/05.

DR N-PSDB; AAF27729.

XX Isolated polypeptide with a human transport protein sequence is useful

PT for the diagnosis, prevention and treatment of disorders associated

XX with the immune, reproductive and cardiovascular systems -

XX

PS Claim 2; Page 130; 165pp; English.

XX

CC The present invention provides the protein and coding sequences for 43

CC novel human transport proteins (designated TPPTs). These can be used in

CC the diagnosis and treatment of transport, metabolic, neurological,

CC reproductive, cardiovascular and immune disorders, and cell proliferative

CC disorders such as cancer.

XX

XX

SQ Sequence 219 AA;

Query Match 100.0%; Score 1178; DB 22; Length 219;

Best Local Similarity 100.0%; Pred. No. 1.5e-85;

Matches 219; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLLWACIVCAFAKRRPFPIGEDDDGHLHPSLNIPYGINLPPLPYRPNVTVP 60

DB 1 MKLLWACIVCAFAKRRPFPIGEDDDGHLHPSLNIPYGINLPPLPYRPNVTVP 60

QY 61 SYPGNTYTDGLSPYWLITSGFPVYVHIRGFPLATQLNVPLPGRGFPVPPSRFFSA 120

DB 61 SYPGNTYTDGLSPYWLITSGFPVYVHIRGFPLATQLNVPLPGRGFPVPPSRFFSA 120

QY 121 AAAPAAPPIAAEPAAAPLTATPVAAEPAAAGAPVAAEPAAEPVGAEPAAEPVAAEPAA 180

DB 121 AAAPAAPPIAAEPAAAPLTATPVAAEPAAAGAPVAAEPAAEPVGAEPAAEPVAAEPAA 180

QY 181 EAPVGEPAEPEPSAEPATAKPAAPPEHPSPSLEQANQ 219

DB 181 EAPVGEPAEPEPSAEPATAKPAAPPEHPSPSLEQANQ 219

RESULT 2

AAU74536

ID AAU74536 standard; Protein; 219 AA.

XX

AC AAU74536;

XX

XX

DT 23-APR-2002 (first entry)

XX

DE Human zsig63 polypeptide.

XX

XX Human; zsig63; chromosome 4q12-4q13; salivary protein; antimicrobial;

KW microbial infection; tooth decay; periodontal disease; thrush; emphysema;

KW gastrointestinal disease; urinary tract infection; vaginal infection;

KW skin infection; epithelial wound; chronic tissue damage; cystic fibrosis;

KW acquired immunodeficiency syndrome; AIDS; lung infection; sarcoidosis;

KW chronic bronchitis; gene therapy; protein therapy.

XX

XX Homo sapiens.

OS

XX

XX US6331413-B1.

PN

XX

XX 18-DEC-2001.

PD

XX

XX 17-MAR-2000; 2000US-0527345.

PF

XX

XX 17-MAR-1999; 99US-124820P.

PR

XX

XX (ZYMO) ZYMOGENETICS INC.

PA

XX

XX Adler DA, Sheppard PO;

PI

XX

XX WPI; 2002-096707/13.

DR

XX

XX N-PSDB; AAS20591.

DR

XX

XX Polynucleotides encoding salivary proteins useful as anti-microbial

PT agents -

XX

PS

XX Claim 1; Column 49-52; 29pp; English.

XX

CC The invention relates to a polynucleotide derived from the 4q12-4q13

CC region of human chromosome 4 and encoding a zsig63 polypeptide, a

CC secreted salivary protein with anti-microbial activity. Due to their

CC microbial activity, the sequences can be used in the study of microbial

CC infections, e.g. for recombinant production of anti-microbial proteins.

CC The sequences can be used in the treatment of tooth decay, periodontal

CC disease, thrush, gastrointestinal disease, urinary tract infections,

CC vaginal infections, skin infections, epithelial wounds, chronic tissue

CC damage, acquired immunodeficiency syndrome (AIDS), cystic fibrosis, lung

CC infections, sarcoidosis, emphysema and chronic bronchitis. This sequence

CC represents human zsig63.

XX

SQ Sequence 219 AA;

Query Match 100.0%; Score 1178; DB 23; Length 219;

Best Local Similarity 100.0%; Pred. No. 1.5e-85;

Matches 219; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLLWACIVCAFAKRRPFPIGEDDDGHLHPSLNIPYGINLPPLPYRPNVTVP 60

DB 1 MKLLWACIVCAFAKRRPFPIGEDDDGHLHPSLNIPYGINLPPLPYRPNVTVP 60

QY 61 SYPGNTYTDGLSPYWLITSGFPVYVHIRGFPLATQLNVPLPGRGFPVPPSRFFSA 120

DB 61 SYPGNTYTDGLSPYWLITSGFPVYVHIRGFPLATQLNVPLPGRGFPVPPSRFFSA 120

QY 121 AAAPAAPPIAAEPAAAPLTATPVAAEPAAAGAPVAAEPAAEPVGAEPAAEPVAAEPAA 180

DB 121 AAAPAAPPIAAEPAAAPLTATPVAAEPAAAGAPVAAEPAAEPVGAEPAAEPVAAEPAA 180

QY 181 EAPVGEPAEPEPSAEPATAKPAAPPEHPSPSLEQANQ 219

DB 181 EAPVGEPAEPEPSAEPATAKPAAPPEHPSPSLEQANQ 219

RESULT 3

AAW30653

ID AAW30653 standard; Protein; 219 AA.

XX

AC AAW30653;

XX

XX

DT 12-APR-1999 (first entry)

XX

XX Human secreted protein clone cp116 1 protein.

DE

XX Human; secreted protein; nutritional activity; cytokine; vaccine;

KW cell proliferation; differentiation; immune stimulation; suppression;

KW haematopoiesis regulation; tissue growth; activin; inhibin; chemotactic;

KW chemokinetic; haemostatic; thrombolytic; anti-inflammatory; gene therapy;

KW tumour invasion suppression; tumour inhibition.

XX

XX Homo sapiens.

OS

XX

XX WO9901466-A1.

PN

XX

XX 14-JAN-1999.

PD

XX

XX 01-JUL-1998; 98WO-US13813.

PF

XX

XX 27-OCT-1997; 97US-0958304.

PR

XX

XX 02-JUL-1997; 97US-0887195.

PR

XX

XX (GEMY) GENETICS INST INC.

PA

XX

XX Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Racie LA;

PI

XX

XX Spaulding V, Treacy M;

PI

XX

XX WPI; 1999-105994/09.

DR

XX

XX N-PSDB; AAV80740.

DR

XX

PT New polynucleotides encoding secreted human proteins - are derived
PT from human foetal brain, adult testes, adult brain, foetal kidney,
PT adult salivary glands, or adult blood cDNA libraries, useful as, e.g.
PT potential vaccines
XX
PS Claim 24; Page 71-72; 107pp; English.
XX
CC The present sequence represents a human secreted protein from clone
CC cpl16.1, deposited as ATCC 98482. Human secreted protein clone
CC polynucleotides and proteins are predicted to have biological
CC activities which would make them suitable for treating, preventing or
CC ameliorating medical conditions in humans and animals. Suggested
CC activities include nutritional activity, cytokine and cell
CC proliferation/differentiation activity, immune stimulating (e.g. as
CC vaccines) or suppressing activity, haematopoiesis regulating activity,
CC tissue growth activity, activin/inhibin activity, chemotactic/
CC chemokinetic activity, haemostatic and thrombolytic activity, receptor/
CC ligand activity, anti-inflammatory activity, cadherin/tumour invasion
CC suppressor activity, and tumour inhibition activity. The polynucleotides
CC are also stated to be useful for gene therapy.
XX
SQ Sequence 219 AA;

Query Match 99.5%; Score 1172; DB 20; Length 219;
Best Local Similarity 99.1%; Pred. No. 4.4e-85;
Matches 217; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
QY 1 MKLLWACIVCVAFARKRRFFIGEDNDGHPHLSINIPYGINLPPLPYRPNVTVP 60
Db 1 MKLLWACIVCVAFARKRRFFIGEDNDGHPHLSINIPYGINLPPLPYRPNVTVP 60
QY 61 SYPGNTYTDGLPSYPWLTSPGFPYVYHIRGFPLATQNLNVPPLPRGFFVPPSRFFSA 120
Db 61 SYPGNTYTDGLPSYPWLTSPGFPYVYHIRGFPLATQNLNVPPLPRGFFVPPSRFFSA 120
QY 121 AAAPAAPPIAAEPAAPAAAPLTATPVAAEPAAAGAPVAAEPAEAPVAAEPAEPA 180
Db 121 AAAPAAPPIAAEPAAPAAAPLTATPVAAEPAAAGAPVAAEPAEAPVAAEPAEPA 180
QY 181 EAPVGEPAEAPSPAEPAATAPKPAAPPEPHSPSLEQANQ 219
Db 181 EAPVGEPAEAPSPAEPAATAPKPAAPPEPHSPSLEQANQ 219

RESULT 4
AA19472
ID AA19472 standard; Protein; 221 AA.

XX AA19472;
XX
XX
DT 14-JUL-1999 (first entry)
XX
DE Amino acid sequence of a human secreted protein.
KW Human secreted protein; cancer; tumour; neurodegenerative disorder;
KW developmental abnormality; fetal deficiency; blood disorder; leukemia;
KW immune system disease; autoimmune disease; hepatic disease;
KW renal disease; lymphoma; inflammation; allergy; ischemic shock;
KW Alzheimer's; cognitive disorder; schizophrenia; arthritis; prostate disease;
KW obesity; osteoclast; osteoporosis; arthritis; malignancy; testes disease;
KW lung disease; thymus disease; digestive disorder; endocrine disorder;
KW infection; AIDS.
XX
OS Homo sapiens.
XX
PN WO9922243-A1.
XX
PD 06-MAY-1999.
XX
PF 23-OCT-1998; 98WO-US22376.
XX
PR 24-OCT-1997; 97US-0063387.
PR 24-OCT-1997; 97US-0062784.

PR 24-OCT-1997; 97US-0063088.
PR 24-OCT-1997; 97US-0063089.
PR 24-OCT-1997; 97US-0063090.
PR 24-OCT-1997; 97US-0063091.
PR 24-OCT-1997; 97US-0063092.
PR 24-OCT-1997; 97US-0063097.
PR 24-OCT-1997; 97US-0063098.
PR 24-OCT-1997; 97US-0063099.
PR 24-OCT-1997; 97US-0063100.
PR 24-OCT-1997; 97US-0063101.
PR 24-OCT-1997; 97US-0063109.
PR 24-OCT-1997; 97US-0063110.
PR 24-OCT-1997; 97US-0063111.
PR 24-OCT-1997; 97US-0063148.
PR 24-OCT-1997; 97US-0063386.
XX (HUMA-) HUMAN GENOME SCI INC.
XX
PI Brewer LA, Carter KC, Duan DR, Ebner R, Endress GA;
PI Feng P, Florence C, Florence KA, Greene JM, Janat F;
PI Kay H, Lafleur DW, Moore PA, Ni J, Olsen HS, Rosen CA;
PI Ruben SM, Shi Y, Soppet DR, Wei Y, Young P;
XX
XX WPI; 1999-303069/25.
DR N-PSDB; AAX61352.
DR
XX New isolated human genes and the secreted polypeptides they encode
PT
XX Claim 11; Page 401-402; 546pp; English.
XX
CC The specification describes cDNA sequences (AAX61322-X61470) encoding
CC human secreted proteins (AA19442-19590). The polynucleotides and their
CC corresponding secreted polypeptides are useful for preventing, treating
CC or ameliorating medical conditions, e.g. by protein or gene therapy.
CC Pathological conditions can also be diagnosed by determining the amount
CC of the polypeptides in a sample or by determining the presence of
CC mutations in the polynucleotides. Specific uses are described for each
CC of the polynucleotides, based on which tissues they are most highly
CC expressed in, and include developing products for the diagnosis or
CC treatment of cancer, tumours, neurodegenerative disorders, developmental
CC abnormalities and fetal deficiencies, blood disorders, leukemias,
CC diseases of the immune system, autoimmune diseases, hepatic and renal
CC disease, lymphomas, inflammation, allergies, ischemic shock, Alzheimer's
CC and cognitive disorders, schizophrenia, prostate diseases, obesity,
CC disorders involving osteoclasts such as osteoporosis, arthritis or
CC malignancies, diseases of testes, lung or thymus, digestive/endocrine
CC disorders, infections and AIDS. The polypeptides are also useful for
CC identifying their binding partners.
XX
SQ Sequence 221 AA;

Query Match 94.5%; Score 1113.5; DB 20; Length 221;
Best Local Similarity 95.0%; Pred. No. 1.8e-80;
Matches 209; Conservative 1; Mismatches 9; Indels 1; Gaps 1;
QY 1 MKLLWACIVCVAFARKRRFFIGEDNDGHPHLSINIPYGINLPPLPYRPNVTVP 60
Db 1 MKLLWACIVCVAFARKRRFFIGEDNDGHPHLSINIPYGINLPPLPYRPNVTVP 60
QY 61 SYPGNTYTDGLPSYPWLTSPGFPYVYHIRGFPLATQNLNVPPLPRGFFVPPSRFFSA 120
Db 61 SYPGNTYTDGLPSYPWLTSPGFPYVYHIRGFPLATQNLNVPPLPRGFFVPPSRFFSA 120
QY 121 AAAPAAPPIAAEPAAPAAAPLTATPVAAEPAAAGAPVAAEPAEAPVAAEPAEPA 179
Db 121 AAAPAAPPIAAEPAAPAAAPLTATPVAAEPAAAGAPVAAEPAAAGAPVAAEPAEPA 180
QY 180 EAPVGEPAEAPSPAEPAATAPKPAAPPEPHSPSLEQANQ 219
Db 181 EAPVGEPAEAPSPAEPAATAPKPAAPPEPHSPSLEQANQ 220

RESULT 5

Db 88 --PHAPPQALPPPPALPPPPP--ASPPLLPPALSPPLPAPPPPSAPPASPPPPAPPPPPAP 143

AAW06915
in AAW06915 standard. Protein: 207 AA

```
XX AAW06915;
AC
XX 02-APR-1997 (first entry)
DT
XX T. cruzi TcD antigen.
DE
XX TcD; diagnosis; infection; antigen; Chagas disease.
KW
XX Trypanosoma cruzi.
OS
XX
XX Key Location/Qualifiers
FH 6..20
FT /label= Antigenic_epitope
FT /note= "Claim 3, page 38"
XX
XX WO9629605-A2.
PN
XX 26-SEP-1996.
PD
XX
XX 12-MAR-1996; 96WO-US03380.
PF
XX
XX 14-MAR-1995; 95US-0403379.
PR
XX (CORI-) CORIXA CORP.
PA
XX Reed SG;
PI
XX WPI; 1996-485445/48.
DR
XX N-PSDB; AAT46149.
DR
XX
XX Detecting and preventing T. cruzi infection - using polypeptide(s)
PT or antibodies contg. or reactive with antigen epitope(s) of T. cruzi
PT proteins
XX
XX Disclosure; Fig 4; 59pp; English.
PS
XX The TcD antigen (AAW06915) of Trypanosoma cruzi includes an
CC antigenic epitope (see also AAW06917) that can be utilised, pref.
CC with other T. cruzi epitopes (see also AAW06914-19), in immunoassays
CC for T. cruzi infection (Chagas' disease) and in vaccine compns.
XX
XX Sequence 207 AA;
SQ
Query Match 15.2%; Score 179; DB 17; Length 207;
Best Local Similarity 40.8%; Pred. No. 1.1e-06;
Matches 42; Conservative 11; Mismatches 46; Indels 4; Gaps 2;
QY 114 PSRFFSAAAPAPPTAAEPAAAPLTATPVAAEPAAAGAPVAAEPAAAPVGAEPAAEAP 173
Db 105 PAESKSAEPKPAEPK-SAEPPKPAEPKSAEPKPAEPKSAEPKSAEPKSAEPKSAAGP 163
QY 174 VAAEPAAEPVGVPEAAEPSPAEPAETAKPAAPPEP---HPSPS 213
Db 164 KPAEPKSAEPKPAEPKSAEPKSAEPKSAEPKSAEPKSAEPKSAEPKPA 206
RESULT 8
AAV32850
ID AAY32850 standard; Protein; 207 AA.
XX
XX AAY32850;
AC
XX 29-OCT-1999 (first entry)
DT
XX TcD protein sequence.
DE
XX TcD repeat sequence; 35kd T. cruzi homologue; immune system disorder;
KW eukaryotic ribosomal protein L19E; infection; detection; diagnosis;
KW Chagas' disease; Leishmania infection.
XX
XX Trypanosoma cruzi.
OS
XX
```

```
PN US5942403-A.
XX
XX 24-AUG-1999.
PD
XX 15-SEP-1997; 97US-0929414.
PF
XX 15-SEP-1997; 97US-0929414.
PR
XX 14-MAR-1995; 95US-0403379.
XX
XX (CORI-) CORIXA CORP.
PA
XX Houghton R, Reed SG, Skeiky YAW;
XX WPI; 1999-517419/43.
XX N-PSDB; AAZ10979.
XX
XX Trypanosoma cruzi antigenic homologue of eukaryotic ribosomal
PT protein L19E, useful for screening assays to detect T. cruzi
PT infection
XX
XX Example 3; Fig 4; 26pp; English.
PS
XX
XX This sequence is the Trypanosoma cruzi TcD protein. Antigenic fragments
CC of TcD and fragments of the 35 kD Trypanosoma cruzi homologue (TcE) of
CC the eukaryotic ribosomal protein L19E can be used in the method of the
CC invention for detecting T. cruzi infection in a biological sample, which
CC comprises: (a) contacting a biological sample with a first polypeptide
CC comprising the sequence shown in AAY32839; (b) contacting the biological
CC sample with a second polypeptide comprising the TcD sequence shown in
CC AAY32840 or AAY32841; and (c) detecting the presence of antibodies that
CC bind to at least one of the polypeptides, indicating T. cruzi infection.
CC The methods and polypeptides may be used to detect T. cruzi (which
CC causes Chagas' disease) and a variety of immune system disorders) and/or
CC Leishmania infection in individuals and blood supplies. The compounds
CC and methods may also be used to protect against T. cruzi infection. The
CC N-terminal region of TcE (residues 1-136) cross-reacts with
CC anti-Leishmania antibodies and can be specifically excluded from the
CC antigenic polypeptide to avoid such cross-reactivity.
XX
XX Sequence 207 AA;
SQ
Query Match 15.2%; Score 179; DB 20; Length 207;
Best Local Similarity 40.8%; Pred. No. 1.1e-06;
Matches 42; Conservative 11; Mismatches 46; Indels 4; Gaps 2;
QY 114 PSRFFSAAAPAPPTAAEPAAAPLTATPVAAEPAAAGAPVAAEPAAAPVGAEPAAEAP 173
Db 105 PAESKSAEPKPAEPK-SAEPPKPAEPKSAEPKPAEPKSAEPKSAEPKSAEPKSAAGP 163
QY 174 VAAEPAAEPVGVPEAAEPSPAEPAETAKPAAPPEP---HPSPS 213
Db 164 KPAEPKSAEPKPAEPKSAEPKSAEPKSAEPKSAEPKSAEPKSAEPKPA 206
RESULT 9
AAV23305
ID AAY23305 standard; Peptide; 207 AA.
XX
XX AAY23305;
AC
XX 31-AUG-1999 (first entry)
DT
XX TcD antigen of Trypanosoma cruzi.
DE
XX Trypanosoma cruzi epitope; Trypanosoma cruzi infection; antigen;
KW vaccine; Chagas' disease; TcD antigen.
XX
XX Trypanosoma cruzi.
OS
XX WO9931246-A1.
PN
XX 24-JUN-1999.
PD
XX
```

PF 04-DEC-1998; 98WO-US25871.
XX
PR 18-DEC-1997; 97US-0993674.
XX
XX (CORI-) CORIXA CORP.
XX
PI Houghton RL, Lodes MJ, McNeill PD, Reed SG, Skeiky YAW;
PI Smith JM;
XX
DR WPI; 1999-405035/34.
XX
XX New isolated Trypanosoma cruzi epitopes
PT
XX Disclosure; Page 83-84; 103pp; English.
PS
XX The specification describes new Trypanosoma cruzi epitopes. A method
CC for detecting Trypanosoma cruzi infection in a biological sample
CC comprises contacting the sample with a polypeptide comprising
CC an epitope of a TC antigen, or a variant of the antigen that differs
CC only in conservative substitutions and/or modifications and detecting
CC the presence of antibodies that bind to the polypeptide in the sample,
CC thereby detecting TC infection. The TC polypeptides can be used in
CC vaccines for inducing protective immunity against Chagas' disease in
CC a patient. The polypeptides and antibodies can also be used for detecting
CC TC infection. The present sequence represent a TcD antigen, which
CC can be used in the above assay to improve sensitivity.
XX
XX Sequence 207 AA;
SQ
Query Match 15.2%; Score 179; DB 20; Length 207;
Best Local Similarity 40.8%; Pred. No. 1.1e-06;
Matches 42; Conservative 11; Mismatches 46; Indels 4; Gaps 2;
QY 114 PSRFFSAAAPAPPTAEPAPAAAPLTATPVAAPAAAGAPVAAAPVCAEPAAAP 173
DB 105 PAESKSAEPKPAEPK-SAEPPKPAEPKSAEPKPAEPKSAEPKPAEPKSAEPK 163
QY 174 VAAEPAAAPVGVPEAAEPSPAEPAATKAPAAPEP---HPSPS 213
DB 164 KPAEPKSAEPKPAEPKSAEPKSAEPKSAEPKSAEPKSAEPKPA 206
RESULT 10
AAM42089
ID AAM42089 standard; Protein; 718 AA.
XX
XX AAM42089;
AC
XX
XX 22-OCT-2001 (first entry)
DT
XX
XX Human polypeptide SEQ ID NO 7020.
DE
XX Human; nontropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia.
XX
XX Homo sapiens.
OS
XX
XX WO200153312-A1.
PN
XX
XX 26-JUL-2001.
PD
XX
XX 26-DEC-2000; 2000WO-US34263.
PF
XX
XX 21-JAN-2000; 2000US-0488725.
PR
XX 25-APR-2000; 2000US-0552317.
PR
XX 05-JUL-2000; 2000US-0598042.
PR
XX 19-JUL-2000; 2000US-0620312.
PR
XX 03-AUG-2000; 2000US-0653450.
PR
XX 14-SEP-2000; 2000US-0662191.

PR 19-OCT-2000; 2000US-0693036.
XX
XX 29-NOV-2000; 2000US-0727344.
XX
XX (HYSE-) HYSEQ INC.
PA
XX
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
XX WPI; 2001-442253/47.
DR
DR N-PSDB; AAI61245.
XX
XX Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
XX Example 2; SEQ ID NO 7020; 10078pp; English.
PS
XX The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC the encoded polypeptides (AAM38642-AAM42213) with nontropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
XX Sequence 718 AA;
SQ
Query Match 14.9%; Score 175.5; DB 22; Length 718;
Best Local Similarity 32.3%; Pred. No. 7.4e-06;
Matches 60; Conservative 16; Mismatches 61; Indels 49; Gaps 9;
QY 35 HPSINIPYGI RNLP---PPLYRPVNTVPSPVPGNTYDTGLPSVPVTLTSPGPPVYVHIR 91
DB 394 HKQVNVFNNTSSPTSNPVTTTKPVTTT-----KPVTTTTTKP-----VTTTTKP----- 437
QY 92 GFPLATQINVPPLPPRGPEFPVPPSRFFSAAAPAA-APPIAAEPAAAPLTATPVAAPAA 150
DB 438 ----VTIINQPSVKP-----AAAKPAPAKPVAAKPVATKTATVRP----- 473
QY 151 GAPVAAEPAAEA-PVGAEPAAEPVAAEPAAEPVGEPAEPSPAEPAATKAPAAPEPH 209
DB 474 --PVAVKPATAAKPAAKPAAVRPAA--AAAKPVATKPEVPRPQAAKPAATKPAATKPM 529
QY 210 PSPSLE 215
DB 530 VKMSRE 535
RESULT 11
ABP41277
ID ABP41277 standard; Protein; 903 AA.
XX
XX ABP41277;
AC
XX
XX 23-AUG-2002 (first entry)
DT
XX
XX Human ovarian antigen HOCQH66, SEQ ID NO:2409.
DE
XX Human; ovarian antigen; ovary; ovarian; breast; cancer; tumour;
KW ovarian cancer; breast cancer; reproductive system disorder;
KW infertility; pregnancy disorder; anovulation; polycystic ovary syndrome;
KW PCOS; ovarian cyst; dysmenorrhoea; endocrine disorder; infection;
KW inflammatory condition; immune disorder; blood disorder;
KW cardiovascular disorder; respiratory disorder; neurological disorder;


```
XX SQ Sequence 3177 AA;
Query Match 14.9%; Score 175.5; DB 22; Length 3177;
Best Local Similarity 32.3%; Pred. No. 3.5e-05;
Matches 60; Conservative 16; Mismatches 61; Indels 49; Gaps 9;
QY 35 HPSLNIPYGIKRLP---PPLYYRPVNTVPSYPGNTYDTGLPSYPWILTSPGFPVYVHIR 91
Db 2853 HKQVNVNNTSSPTSNPVTTKPVTTT-----KPVITTTKP-----VTTTTPK----- 2896
QY 92 GFPLATQNLNVPPLPRGPFVPPSFRFSAARA-APPIAAEPAAPAAAPLTATPVAAEPAA 150
Db 2897 ----VTIINQPSVKP-----AAKAPAKPVAKPVATKTATVRP----- 2932
QY 151 GAPVAEPAEAE-PVGAEPAAEAPVAEPAEAPYGVFEPAEAEPSPAEPATAKPAAPPEPH 209
Db 2933 --PVAVKPATAKPVAKPAAPVPPAA--AAAKPVATKPEVPRPQAAKPAATKPAATTKPM 2988
QY 210 PPSLE 215
Db 2989 VKMSRE 2994

RESULT 13
ABB61040
ID ABB61040 standard; Protein; 566 AA.
AC ABB61040;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 9912.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE ) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
PI WPI; 2001-656860/75.
XX
DR N-PSDB; ABL05143.
XX
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions .
XX
XX Disclosure; SEQ ID NO 9912; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
```

```
SQ Sequence 566 AA;
Query Match 14.8%; Score 174; DB 22; Length 566;
Best Local Similarity 30.2%; Pred. No. 7.6e-06;
Matches 77; Conservative 13; Mismatches 107; Indels 58; Gaps 7;
QY 17 KRRFPFTIGEDNDGDHPLHPSLNIPYGIKRLNPPPLYYRPVNT----- 58
Db 312 KERYCMIGDSLDAFVDLIKGLPEFNNPRNPKPTTKLPTTPEELAAMEEAAAAA 371
QY 59 -----VPSYPGNTYDTGLPSYPWILT---SPGFPYVYHI-----R 91
Db 372 AAEAAEAAAAAGEAGPDGAPAAEGEAKAPAKEPTTPEPTPPPPPPFFEYSIDLPE 431
QY 92 GFPLATQNLNVPPLPRGPFVPPSFRFSAARAAP--AAPPT-AAEPAAPAAAPLTATPVAAEP 148
Db 432 GAEPYVYKNTPEPPPGSEPEFVPAEAGEAAPAAEGRAAPPAEGNVPADGAAPP 491
QY 149 AAGAPVAAEPAEAPVGAEPAAEAPV-----AEPFAEAPVGVFEPAAB-----EPSPAEP 198
Db 492 AEGNAPAAEGNAPADGAAPPAAEAAAPADAAPAAEAPAAEAPAAEATAAEAPPAEAP 551
QY 199 ATAAPAAPEPHPSPS 213
Db 552 AEAAPAAAEAGEAPPA 566

RESULT 14
AAB82806
ID AAB82806 standard; Protein; 538 AA.
XX
AC AAB82806;
XX
DT 12-NOV-2001 (first entry)
XX
DE Human low density lipoprotein binding protein 2 (LBP-2).
XX
KW Low density lipoprotein binding protein 2; LBP-2; LDL; human;
KW atherosclerosis; antiarteriosclerotic; therapy; diagnosis;
KW vaccine.
XX
OS Homo sapiens.
XX
PN WO200164874-A2.
XX
PD 07-SEP-2001.
XX
PF 28-FEB-2001; 2001WO-US06356.
XX
PR 02-MAR-2000; 2000US-0517849.
PR 14-JUL-2000; 2000US-0616289.
XX
XX (BOST-) BOSTON HEART FOUND INC.
XX
XX Lees AM, Lees RS, Law SW, Arjona AA;
XX
XX WPI; 2001-565505/63.
XX
XX N-PSDB; AAB28499.
XX
XX New isolated low density lipoprotein binding polypeptide for treating,
PT diagnosing and/or identifying therapeutic agents for atherosclerosis -
PT Claim 13(j); Fig 7A; 143pp; English.
XX
XX The present sequence is that of novel human low density lipoprotein
CC binding protein 2 (LBP-2). The amino acid sequence was deduced
CC from the coding region of isolated genomic DNA (see AAB28499).
CC It differs from the sequence predicted from an LBP-2 cDNA clone
CC (see AAB82803) by the presence of an additional 321 amino acids
CC at the N-terminus (the cDNA clone is 5' truncated). Human
CC LBP-2 is an example of claimed LBP polypeptides of the invention
CC that are capable of binding to native and methylated low density
CC lipoproteins. Also claimed are biologically active fragments and
```


GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2003, 13:44:29 ; Search time 36 Seconds
(without alignments)
178.989 Million cell updates/sec

Title: US-09-923-236-2
Perfect score: 1178
Sequence: 1 MKLLWACIVCVAFARKRRF.....TAKPAAPHPSPSLEQANQ 219

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA.*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/1/iaa/backfilesi.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1178	100.0	219	4	US-09-527-345-2
2	179	15.2	207	1	US-07-836-642-2
3	179	15.2	207	1	US-08-189-363-2
4	179	15.2	207	1	US-08-403-379A-6
5	179	15.2	207	2	US-08-557-309B-50
6	179	15.2	207	2	US-08-929-414-6
7	179	15.2	207	3	US-08-834-306-50
8	179	15.2	207	4	US-08-993-674A-50
9	179	15.2	207	4	US-09-256-976-50
10	170	14.4	267	4	US-08-112-142
11	170	14.4	267	4	US-08-818-111-137
12	170	14.4	267	4	US-09-056-556-142
13	170	14.4	267	4	US-09-072-596-137
14	163.5	13.9	214	1	US-08-217-327-4
15	154	13.1	805	4	US-09-103-429A-4
16	153	13.0	786	4	US-09-103-429A-3
17	148	12.6	2972	4	US-09-579-181-2
18	148	12.6	3118	4	US-09-579-181-1
19	145.5	12.4	1565	6	5352450-2
20	139.5	11.8	941	4	US-07-757-022B-14
21	139.5	11.8	1022	4	US-07-757-022B-84
22	139.5	11.8	1038	4	US-07-757-022B-74
23	139.5	11.8	1049	4	US-07-757-022B-58
24	139.5	11.8	1140	4	US-07-757-022B-104
25	139.5	11.8	1270	4	US-07-757-022B-44
26	139.5	11.8	1311	4	US-07-757-022B-44
27	139.5	11.8	1313	4	US-07-757-022B-142

28	139.5	11.8	1314	4	US-07-757-022B-50	Sequence 50, Appl
29	139.5	11.8	1320	4	US-07-757-022B-46	Sequence 46, Appl
30	139.5	11.8	1320	4	US-07-757-022B-60	Sequence 60, Appl
31	139.5	11.8	1354	4	US-07-757-022B-48	Sequence 48, Appl
32	139.5	11.8	1361	4	US-07-757-022B-40	Sequence 40, Appl
33	139.5	11.8	1363	4	US-07-757-022B-52	Sequence 52, Appl
34	139.5	11.8	1404	4	US-07-757-022B-2	Sequence 2, Appl
35	139.5	11.8	1404	4	US-07-757-022B-62	Sequence 62, Appl
36	138	11.7	180	6	5273901-7	Patent No. 5273901
37	138	11.7	180	6	5482709-6	Patent No. 5482709
38	136.5	11.6	504	4	US-09-219-849-3	Sequence 3, Appl
39	136.5	11.6	561	1	US-08-642-255-52	Sequence 52, Appl
40	135	11.5	1274	4	US-09-095-443-2	Sequence 2, Appl
41	133.5	11.3	234	1	US-08-642-255-51	Sequence 51, Appl
42	133.5	11.3	1185	4	US-09-041-886-23	Sequence 23, Appl
43	132	11.2	267	3	US-08-301-162-16	Sequence 16, Appl
44	132	11.2	267	4	US-09-461-240-16	Sequence 16, Appl
45	132	11.2	267	4	US-09-968-927-16	Sequence 16, Appl

ALIGNMENTS

RESULT 1
US-09-527-345-2
; Sequence 2, Application US/09527345
; Patent No. 6331413
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; TITLE OF INVENTION: SECRETED SALIVARY ZSIG63 POLYPEPTIDE
; FILE REFERENCE: 97-71
; CURRENT APPLICATION NUMBER: US/09/527,345
; CURRENT FILING DATE: 1999-03-17
; PRIOR APPLICATION NUMBER: US 60/124,820
; PRIOR FILING DATE: 1999-03-17
; NUMBER OF SEQ. ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ. ID NO. 2
; LENGTH: 219
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-527-345-2

Query Match	100.0%	Score 1178;	DB 4;	Length 219;
Best Local Similarity	100.0%;	Pred. No. 2.5e-93;		
Matches 219;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	MKLLWACIVCVAFARKRRFPFIFGEDDDGHPHPSLNIPYGI RNLPPLPYRPVNTVP	60	
Db	1	MKLLWACIVCVAFARKRRFPFIFGEDDDGHPHPSLNIPYGI RNLPPLPYRPVNTVP	60	
QY	61	SYPGNTYTDGLSPYWILTSFGFVYVYHIRGFPLATQNLNVPPLPGRGFFVPPSRFFSA	120	
Db	61	SYPGNTYTDGLSPYWILTSFGFVYVYHIRGFPLATQNLNVPPLPGRGFFVPPSRFFSA	120	
QY	121	AAAPAAPTAAPAAAPLITATPVAAEPAAAPVAAEPAAAPVAAEPAAAPVAAEPAA	180	
Db	121	AAAPAAPTAAPAAAPLITATPVAAEPAAAPVAAEPAAAPVAAEPAAAPVAAEPAA	180	
QY	181	EAPVGVEPAEPPSPAPATAKPAPEPHPSLEQANQ	219	
Db	181	EAPVGVEPAEPPSPAPATAKPAPEPHPSLEQANQ	219	
RESULT 2				
US-07-836-642-2				
; Sequence 2, Application US/07836642				
; Patent No. 5304371				
; GENERAL INFORMATION:				
; APPLICANT: Reed, Steven G.				
; TITLE OF INVENTION: Peptide for Diagnosing and Immunizing				
; TITLE OF INVENTION: Against T. cruzi Infection				

```
;
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Immunex Corporation
; STREET: 51 University street
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98101
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/836,642
; FILING DATE: 14-FEB-1992
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Perkins, Patricia A.
; REGISTRATION NUMBER: 34,693
; REFERENCE/DOCKET NUMBER: REED
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206)587-0430
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 207 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-836-642-2

Query Match 15.2%; Score 179; DB 1; Length 207;
Best Local Similarity 40.8%; Pred. No. 3.2e-08;
Matches 42; Conservative 11; Mismatches 46; Indels 4; Gaps 2;

QY 114 PSRFFSAAAPPTAEPAAAPLTATPVAAEPAAAGAPVAAEPAAEPVGAEPAAEAP 173
|: || || || || || || || || || || || || || || || || || || ||
Db 105 PAESKSAEPKPAEPK-SAEPKPAEPKSAEPKPAEPKSAEPKPAEPKSAEPKSAEPK 163
QY 174 VAAEPAAEPVGEPAEPSPAEPAATAPKPAEP-HPSPS 213
|||: || || || || || || || || || || || || || || || || || ||
Db 164 KPAEPKSAEPKPAEPKSAEPKSAEPKSAEPKSAEPKSAEPKSAEPKPA 206
|: || || || || || || || || || || || || || || || || || || ||

RESULT 3
US-08-169-563-2
; Sequence 2, Application US/08169563
; Patent No. 5413912
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; TITLE OF INVENTION: Peptide for Diagnosing and Immunizing
; TITLE OF INVENTION: Against T. cruzi Infection
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Immunex Corporation
; STREET: 51 University street
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98101
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Apple Operating System 7.1
; SOFTWARE: Microsoft Word For Apple, Version 5.1a
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/169,563
; FILING DATE: 17-DEC-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION NUMBER:
; APPLICATION NUMBER: US 07/836,642
; FILING DATE: 14-FEB-1992
; CLASSIFICATION: 435
```

```
;
; ATTORNEY/AGENT INFORMATION:
; NAME: Perkins, Patricia A.
; REGISTRATION NUMBER: 34,693
; REFERENCE/DOCKET NUMBER: REED-A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206)587-0430
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 207 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-169-563-2

Query Match 15.2%; Score 179; DB 1; Length 207;
Best Local Similarity 40.8%; Pred. No. 3.2e-08;
Matches 42; Conservative 11; Mismatches 46; Indels 4; Gaps 2;

QY 114 PSRFFSAAAPPTAEPAAAPLTATPVAAEPAAAGAPVAAEPAAEPVGAEPAAEAP 173
|: || || || || || || || || || || || || || || || || || || ||
Db 105 PAESKSAEPKPAEPK-SAEPKPAEPKSAEPKPAEPKSAEPKPAEPKSAEPKSAEPK 163
QY 174 VAAEPAAEPVGEPAEPSPAEPAATAPKPAEP-HPSPS 213
|||: || || || || || || || || || || || || || || || || || ||
Db 164 KPAEPKSAEPKPAEPKSAEPKSAEPKSAEPKSAEPKSAEPKSAEPKPA 206
|: || || || || || || || || || || || || || || || || || || ||

RESULT 4
US-08-403-379A-6
; Sequence 6, Application US/08403379A
; Patent No. 5756662
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DETECTION
; TITLE OF INVENTION: OF T. CRUZI INFECTION
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/403,379A
; FILING DATE: 14-MAR-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.406
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 207 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-403-379A-6

Query Match 15.2%; Score 179; DB 1; Length 207;
Best Local Similarity 40.8%; Pred. No. 3.2e-08;
Matches 42; Conservative 11; Mismatches 46; Indels 4; Gaps 2;

QY 114 PSRFFSAAAPPTAEPAAAPLTATPVAAEPAAAGAPVAAEPAAEPVGAEPAAEAP 173
|: || || || || || || || || || || || || || || || || || || ||
```


STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-056-556-142

Query Match 14.4%; Score 170; DB 4; Length 267;
Best Local Similarity 28.0%; Pred. No. 2.5e-07;
Matches 56; Conservative 14; Mismatches 94; Indels 36; Gaps 11;

QY 21 PFIGEDDDGHLHPSLNIPY--GIRNLPPPLYRPNV--TVPSYPGNTYTDGLPSYPW 77
Db 79 PRGTSFGVGGTASPAPEAPVGVVPAVPI---PVPILIPPPG---WQPGMTIP- 131
QY 78 ILTSPGFPVYVHGRPLATQLNVPLPRGFFVPPSRFFSAAAPAPPIAAEPAAPAAA 137
Db 132 --TAP--PTT-----PVTTSATPTPTPTPTPT-----PTTTP--TTPTVTP 170
QY 138 PLT--ATPVAEPAAGAPVAEPAEAPVGAEPAAEAPVAEPAEAPVGVPEAAEAPS- 194
Db 171 PTPPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 230
QY 195 --PAEPATAKPAAPPEHPSP 212
Db 231 QMPTQQQTVAPOTVAPAPQP 250

RESULT 13

US-09-072-596-137
; Sequence 137, Application US/09072596
; Patent No. 6458366
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 350
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 137:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 267 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-072-596-137

Query Match 14.4%; Score 170; DB 4; Length 267;
Best Local Similarity 28.0%; Pred. No. 2.5e-07;
Matches 56; Conservative 14; Mismatches 94; Indels 36; Gaps 11;

QY 21 PFIGEDDDGHLHPSLNIPY--GIRNLPPPLYRPNV--TVPSYPGNTYTDGLPSYPW 77
Db 79 PRGTSFGVGGTASPAPEAPVGVVPAVPI---PVPILIPPPG---WQPGMTIP- 131
QY 78 ILTSPGFPVYVHGRPLATQLNVPLPRGFFVPPSRFFSAAAPAPPIAAEPAAPAAA 137
Db 132 --TAP--PTT-----PVTTSATPTPTPTPTPT-----PTTTP--TTPTVTP 170
QY 138 PLT--ATPVAEPAAGAPVAEPAEAPVGAEPAAEAPVAEPAEAPVGVPEAAEAPS- 194
Db 171 PTPPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 230
QY 195 --PAEPATAKPAAPPEHPSP 212
Db 231 QMPTQQQTVAPOTVAPAPQP 250

RESULT 14

US-08-217-327-4
; Sequence 4, Application US/08217327
; Patent No. 5474925
; GENERAL INFORMATION:
; APPLICANT: John, Maliyakal E
; APPLICANT: Barton, Kenneth A
; TITLE OF INVENTION: Immobilized Proteins in Cotton Fiber
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles and Brady
; STREET: P.O. Box 2113
; CITY: Madison
; STATE: WI
; COUNTRY: USA
; ZIP: 53701-2113
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/812,233
; FILING DATE: 19-DEC-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Seay, Nicholas J
; REGISTRATION NUMBER: 27,386
; REFERENCE/DOCKET NUMBER: 1122990831
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 608-251-5000
; TELEFAX: 608-251-9166
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 214 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-217-327-4

Query Match 13.9%; Score 163.5; DB 1; Length 214;
Best Local Similarity 30.3%; Pred. No. 7e-07;
Matches 67; Conservative 20; Mismatches 81; Indels 53; Gaps 14;

QY 2 KILLWACIVCVAFARKRRFPFTGGDDNDGHLHPSLNIPYGIKRLMPPPLYRPNVTPS 61
Db 6 KNLFSLALLCIAVA-----GVLGQ-----APSNPTSTP-----ATPTP-----PASTPPP 46


```

Qy 62 YPGNTYTDGLPSYPWILTSFGPPYVYHIGFFLATQLNVDPPL----PPRGFPFVPVPSRF 111
Db 47 TQAPPTPTATP--PPVSTPP-----ETSSP---PPVTASPPVPVSTP--PPS-- 86
Qy 118 FSAAAAPAAPIAAEPAAPAAATATPVAAEPAAGAPVAAEPAAEAPVGAEPAAEAPVAAE 177
Db 87 --SPPATPPASPPATPPASPPATPPASPPATPPATPPATPPATPPATPPATPPA-- 141
Qy 178 PAAEAPVGVPEAAEPPSPAE-PATAKPAAEPIHPSLEQA 217
Db 142 PLASPPATV-PAI---SPVQTPLTSPAPPTAPATLGA 178

RESULT 15
US-09-103-429A-4
; Sequence 4, Application US/09103429A
; Patent No. 6187558
; GENERAL INFORMATION:
; APPLICANT: Granados, Robert R
; APPLICANT: Wang, Ping
; TITLE OF INVENTION: A No. 6187558el Invertebrate Intestinal Mucin
; TITLE OF INVENTION: cDNA and Related Products and Methods
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brown, Pinnisi & Michaels, P.C.
; STREET: 118 No. 6187558th Tioga
; CITY: Ithaca
; STATE: NY
; COUNTRY: USA
; ZIP: 14850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/103,429A
; FILING DATE: 24-JUN-1998
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Michaels, Christopher A
; REGISTRATION NUMBER: 34,390
; REFERENCE/DOCKET NUMBER: BTI-39
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (607) 256-2000
; TELEFAX: (607) 256-3628
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 805 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEetical: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Trichoplusia ni
; TISSUE TYPE: peritrophic membrane
US-09-103-429A-4

```


GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2003, 13:47:44 ; Search time 75 Seconds
(without alignments)
74.603 Million cell updates/sec

Title: US-09-923-236-2
Perfect score: 1178
Sequence: 1 MKLLWACIVCAFAKRRF.....TAKPAPEHPSPSLEQANQ 219

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 140259 seqs, 25548876 residues
Total number of hits satisfying chosen parameters: 140259

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA: *
1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep: *
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep: *
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep: *
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep: *
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep: *
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep: *
7: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep: *
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep: *
9: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep: *
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep: *
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep: *
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep: *
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep: *
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1178	100.0	219	9	US-09-922-469-2
2	1178	100.0	219	10	US-09-922-480-2
3	1178	100.0	219	10	US-09-923-236-2
4	173	14.7	538	9	US-09-976-740-43
5	173	14.7	538	12	US-10-023-529-43
6	173	14.7	538	12	US-10-023-523-43
7	170	14.4	428	9	US-09-906-514-4
8	169.5	14.4	802	10	US-09-823-240-2
9	168.5	14.3	550	9	US-09-976-740-47
10	168.5	14.3	550	12	US-10-023-529-47
11	168.5	14.3	550	12	US-10-023-523-47
12	159.5	13.5	537	10	US-09-888-615-104
13	154.5	13.1	433	9	US-09-906-514-2
14	145	12.3	1004	9	US-09-738-626-5676
15	143	12.1	666	10	US-09-791-171-70
16	142.5	12.1	635	9	US-09-738-626-6614
17	142.5	12.1	1023	9	US-09-893-519A-14
18	140	11.9	562	10	US-09-879-792-12
19	139.5	11.8	941	12	US-10-124-557-14

20	139.5	11.8	1022	12	US-10-124-557-84	Sequence 84, Appl
21	139.5	11.8	1038	12	US-10-124-557-74	Sequence 74, Appl
22	139.5	11.8	1049	12	US-10-124-557-58	Sequence 58, Appl
23	139.5	11.8	1140	12	US-10-124-557-104	Sequence 104, Appl
24	139.5	11.8	1270	12	US-10-124-557-44	Sequence 44, Appl
25	139.5	11.8	1311	12	US-10-124-557-42	Sequence 42, Appl
26	139.5	11.8	1313	12	US-10-124-557-142	Sequence 142, Appl
27	139.5	11.8	1314	12	US-10-124-557-50	Sequence 50, Appl
28	139.5	11.8	1320	12	US-10-124-557-46	Sequence 46, Appl
29	139.5	11.8	1320	12	US-10-124-557-60	Sequence 60, Appl
30	139.5	11.8	1354	12	US-10-124-557-48	Sequence 48, Appl
31	139.5	11.8	1361	12	US-10-124-557-40	Sequence 40, Appl
32	139.5	11.8	1363	12	US-10-124-557-52	Sequence 52, Appl
33	139.5	11.8	1404	12	US-10-124-557-2	Sequence 2, Appl
34	139.5	11.8	1404	12	US-10-124-557-62	Sequence 62, Appl
35	135.5	11.5	731	9	US-10-086-464-17	Sequence 17, Appl
36	135	11.5	455	10	US-09-815-242-11781	Sequence 11781, A
37	135	11.5	1274	9	US-10-020-215-2	Sequence 2, Appl
38	133.5	11.3	162	10	US-09-864-761-43620	Sequence 43620, A
39	132	11.2	681	10	US-09-815-242-11830	Sequence 11830, A
40	131.5	11.2	466	10	US-09-925-301-1355	Sequence 1355, Ap
41	131.5	11.2	559	10	US-09-858-155A-2	Sequence 2, Appl
42	131	11.1	645	9	US-09-964-899-17	Sequence 17, Appl
43	130	11.0	339	10	US-09-961-527A-13	Sequence 13, Appl
44	129.5	11.0	721	9	US-10-086-464-5	Sequence 5, Appl
45	128.5	10.9	373	10	US-09-789-386-6	Sequence 6, Appl

ALIGNMENTS

RESULT 1
US-09-922-469-2
; Sequence 2, Application US/09922469
; Patent No. US20020173027A1
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; APPLICANT: Adler, David A.
; TITLE OF INVENTION: SECRETED SALIVARY ZS1G63 POLYPEPTIDE
; FILE REFERENCE: 97-71
; CURRENT APPLICATION NUMBER: US/09/922,469
; CURRENT FILING DATE: 2001-08-03
; PRIOR APPLICATION NUMBER: US 60/124,820
; PRIOR FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 219
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-922-469-2

Query Match 100.0%; Score 1178; DB 9; Length 219;
Best Local Similarity 100.0%; Pred. No. 4.4e-72;
Matches 219; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MKLLWACIVCAFAKRRFPF	IGEDDNDGHP	LHPSLNIPYGINLP	PPPLYRPNVTVP	60
Db	1	MKLLWACIVCAFAKRRFPF	IGEDDNDGHP	LHPSLNIPYGINLP	PPPLYRPNVTVP	60
Qy	61	SYPGNTYDTGLSPYWIIT	TSFGFVYVHIR	GFPLATQLNV	PPPPPPPSRFFSA	120
Db	61	SYPGNTYDTGLSPYWIIT	TSFGFVYVHIR	GFPLATQLNV	PPPPPPPSRFFSA	120
Qy	121	AAAPAAPPIAAEAAAAP	LTATPVAAEPAAGAP	VAEPAEPAEPAE	PAEPAEPAEPAE	180
Db	121	AAAPAAPPIAAEAAAAP	LTATPVAAEPAAGAP	VAEPAEPAEPAE	PAEPAEPAEPAE	180
Qy	181	EAPVGPAAAEPSPAEP	ATAKPAPEHPSP	SPSLEQANQ	219	
Db	181	EAPVGPAAAEPSPAEP	ATAKPAPEHPSP	SPSLEQANQ	219	

```

Db 61 SYPCNTYDTGLPSYPWILTSFGFVYVHIRGFPLATQLNVPPLPRGRFFVPPPSFFSA 120
QY 121 AAAPAAPPIAAEPAAPAAAPLTATVAAEPAGAPVAAEPAAEPVGAEPAAEPVAAEPAA 180
    |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 121 AAAPAAPPIAAEPAAPAAAPLTATVAAEPAGAPVAAEPAAEPVGAEPAAEPVAAEPAA 180
QY 181 EAPVGVPEAAEPESPAEPATAKPAAPAEPEHPSPSLEQANQ 219
    |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 181 EAPVGVPEAAEPESPAEPATAKPAAPAEPEHPSPSLEQANQ 219

RESULT 4
US-09-976-740-43
; Sequence 43, Application US/09976740
; Publication No. US20020194633A1
; GENERAL INFORMATION:
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
; TITLE OF INVENTION: ATHEROSCLEROSIS
; FILE REFERENCE: 10797-004001
; CURRENT APPLICATION NUMBER: US/09/976,740
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 09/616,289
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 08/979,608
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 60/031,930
; PRIOR FILING DATE: 1996-11-27
; PRIOR APPLICATION NUMBER: US 60/048,547
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 43
; LENGTH: 538
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-976-740-43

Query Match 14.7%; Score 173; DB 9; Length 538;
Best Local Similarity 36.8%; Pred. No. 8.1e-05;
Matches 49; Conservative 7; Mismatches 47; Indels 30; Gaps 5;

QY 102 PPLPRGRGFVPPPSRFFSAAAPAPPIAAEPAAPAAAPLTATVAAEPAGAPVAAEPAAE 161
    || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 111 PPRAPRGAP-----AAAAAAPPPTAPPPTPPVAAAPARAPRAAAAAATAPSP 162
    || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 162 APVGAEPAAE--APVAA--EPAAEAPVGVPE-----AAEPE-----SPAEP 199
    || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 163 GPAQPGPRAQRAAPLAAPPAPPAAPVAPVAPAGPRRRAPPVAAAREPPLPPPPPPAPPQ 222
    || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 200 TAKPAAPPEHPSP 212
    :|| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 223 QQQPPPPPPQPPPP 235
    :|| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 5
US-10-023-529-43
; Sequence 43, Application US/10023529
; Patent No. US20020129388A1
; GENERAL INFORMATION:
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
; TITLE OF INVENTION: ATHEROSCLEROSIS
; FILE REFERENCE: 10797-004001
; CURRENT APPLICATION NUMBER: US/10/023,529


```



```
; LENGTH: 802
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-823-240-2

  Query Match      14.4%; Score 169.5; DB 10; Length 802;
  Best Local Similarity 27.8%; Pred. No. 0.0002;
  Matches 64; Conservative 22; Mismatches 91; Indels 53; Gaps 10;

QY 31 GHPLHSLNIPYGIRNLPPLYR-----PVNTVPS-----YPGNTYTDGLPSYP 76
Db 330 GSAFHVLP--PH-YATVPRPLNKNRSPSVNTSSQPPAAKSCAWPTSNFSLP-PSPP 385
QY 77 WILTSFGFYVHIRGFIPLATQNLNVPPLPR-----GFFVFPSPSRFFSAAAPAA 125
Db 386 IMISSPGKATGRPVLPVCVSPVQMPSPPTAPNGSLDSVTYVPSPPT--SGPAAPP 443
QY 126 APPIAAEPAAAAPLTATPV-----AAEPAAGAPVAAEPAEAPVGAEPAAEAPV 174
Db 444 PPPPPPPPPPPPLPPPLPLPLASLHCGSQASPPPGTGLASTPSKPSVLPSFSAGAPA 503
QY 175 AAEPAAEAPVG---VEPA-----AEPPSPAEPATAKPAPEPHPSPS 213
Db 504 SAETPLNPELGSSASEPGLQAASQPAESPTPQGLVLGPPAPPPPPPLPS 553

RESULT 9
US-09-976-740-47
; Sequence 47, Application US/09976740
; Patent No. US20020194633A1
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
; TITLE OF INVENTION: ATHEROSCLEROSIS
; FILE REFERENCE: 10797-004001
; CURRENT APPLICATION NUMBER: US/09/976,740
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 09/616,289
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 08/979,608
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 60/031,930
; PRIOR FILING DATE: 1996-11-27
; PRIOR APPLICATION NUMBER: US 60/048,547
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 47
; LENGTH: 550
; TYPE: PRT
; ORGANISM: Oryctolagus cuniculus
US-09-976-740-47

  Query Match      14.3%; Score 168.5; DB 9; Length 550;
  Best Local Similarity 38.2%; Pred. No. 0.00017;
  Matches 50; Conservative 6; Mismatches 56; Indels 19; Gaps 3;

QY 102 PPLPPRGFFV-----PPSRFFSAAAAAPPTAAEPAEAPVGAEPAAEPTATPVAEPA 150
Db 112 PPRAPRGGAAPAAAPPTTAPPPPPAPVAAAAAPAPRAAAAAATAAPPSPGPAQGP 171
QY 151 GAPVAAEPAEAPVGAEPAAEAPVAAEPAEAPVGAEPAAEFPSPAEPTAKPAAP---- 206
Db 172 RAQRAAPLAAPPAPAAAPPAAPPAAPRRAPPAAAAVAARESPLPPPP---QPPAPPOQ 228
QY 207 -EPHPSPSLEQ 216
Db 229 QQPPPPPPPPQ 239

RESULT 11
US-10-023-523-47
; Sequence 47, Application US/10023523
; Patent No. US20020152485A1
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
; TITLE OF INVENTION: ATHEROSCLEROSIS
; FILE REFERENCE: 10797-004001
; CURRENT APPLICATION NUMBER: US/10/023,523
; CURRENT FILING DATE: 2001-12-17
; PRIOR APPLICATION NUMBER: US/09/616,289
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 09/517,849
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 08/979,608
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 60/031,930
```


Db 140 SAAKSGAPKPAARAAKPAFSGTPGDAAKKAEPK-AKPGAEPAPR 183

```
RESULT 15
US-09-791-171-70
; Sequence 70, Application US/09791171
; Patent No. US20020094336A1
; GENERAL INFORMATION:
; APPLICANT: ANDERSEN, Peter
; APPLICANT: NIELSEN, Rikke
; APPLICANT: OETTINGER, Thomas
; APPLICANT: RASMUSSEN, Peter Birk
; APPLICANT: ROSENKRANDS, Ida
; APPLICANT: WELDLINGH, Karin
; APPLICANT: FLORIO, Walter
; TITLE OF INVENTION: NUCLEIC ACIDS FRAGMENTS AND POLYPEPTIDE FRAGMENTS
; TITLE OF INVENTION: DERIVED FROM M. TUBERCULOSIS
; FILE REFERENCE: 670001-2002.1
; CURRENT APPLICATION NUMBER: US/09/791,171
; CURRENT FILING DATE: 2001-02-20
; PRIOR APPLICATION NUMBER: 09/050,739
; PRIOR FILING DATE: 1998-03-30
; PRIOR APPLICATION NUMBER: 0376/97
; PRIOR FILING DATE: 1997-04-02
; PRIOR APPLICATION NUMBER: 1277/97
; PRIOR FILING DATE: 1997-11-10
; PRIOR APPLICATION NUMBER: 60/044,624
; PRIOR FILING DATE: 1997-04-18
; PRIOR APPLICATION NUMBER: 60/070,488
; PRIOR FILING DATE: 1998-01-05
; NUMBER OF SEQ ID NOS: 173
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 70
; LENGTH: 666
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-09-791-171-70
```

```
Query Match      12.1%; Score 143; DB 10; Length 666;
Best Local Similarity 29.0%; Pred. No. 0,0099;
Matches 64; Conservative 16; Mismatches 87; Indels 54; Gaps 13;

QY 21 PFIGEDDNDGHPHLSNIPYGINLPPP----LYRPVNTVPSYPGNTYDTGLPSYP 76
   || | : | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 25 PFF--DPSASFPPAPASANLPKPNQTPPTSDDLSERFVSAPPPPPPPPPPTP-MP 81
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 77 WILTSGFYVYHIRGFPLATQLNVPLPPRCGFVFPSPRFTSAAAPAAPI-AAEPAA 135
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 82 IAGEPPSE-----PAASKPPTPMPDIAGPEPAPPK-----PPTPPMPIAGPEP 126
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 136 AAPLTATPVAEPAAAG-APVAAE-----PAAEAPVGA--EPAAEAP----- 173
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 127 APPKPTP--PMPIAGPATPTESQLAPRPPTPTGTGACQOPESPAPHVPSHGHOPR 184
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 174 --VAEPAAEAPVGEPAEESPAEPATAKPAAPPEHPSP 212
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 185 RTAPAPPWAKMPDIGEPPP-----PAPSRP-SASPAEPPTRPAP 220
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

Search completed: February 12, 2003, 13:57:02
Job time : 77 secs


```
Db 1 MKLLWACIVCVAFARKRRFPFIGEDDNDGHLHPSLNIPYGINLPPLYRPNVTVP 60
QY 61 SYPGNTYTDGLSPYWLITSPGFYVYHIRGFPLATQNLNVPPLPRGFFVPPSRFFSA 120
Db 61 SYPGNTYTDGLSPYWLITSPGFYVYHIRGFPLATQNLNVPPLPRGFFVPPSRFFSA 120
QY 121 AAAPAAPPIAAEPAAAPLTATPVAAPAAAGAPVAAEPAAAPVGAEPAAAPVAAEPAA 180
Db 121 AAAPAAPPIAAEPAAAPLTATPVAAPAAAGAPVAAEPAAAPVGAEPAAAPVAAEPAA 180
QY 181 EAPVGVPEAAEPPSPAEPATAKPAAPPEPHSPSLEQANQ 219
Db 181 EAPVGVPEAAEPPSPAEPATAKPAAPPEPHSPSLEQANQ 219

RESULT 2
PCT-US02-08277-723
; Sequence 723, Application PC/TUS0208277
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc.
; TITLE OF INVENTION: Human Secreted Proteins
; FILE REFERENCE: PS90/PCT
; CURRENT APPLICATION NUMBER: PCT/US02/08277
; CURRENT FILING DATE: 2002-03-19
; PRIOR APPLICATION NUMBER: US 60/331,287
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/306,171
; PRIOR FILING DATE: 2001-07-19
; PRIOR APPLICATION NUMBER: US 60/277,340
; PRIOR FILING DATE: 2001-03-21
; NUMBER OF SEQ ID NOS: 1357
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 723
; LENGTH: 219
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US02-08277-723

Query Match 100.0%; Score 1178; DB 1; Length 219;
Best Local Similarity 100.0%; Pred. No. 5.7e-76;
Matches 219; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLLWACIVCVAFARKRRFPFIGEDDNDGHLHPSLNIPYGINLPPLYRPNVTVP 60
Db 1 MKLLWACIVCVAFARKRRFPFIGEDDNDGHLHPSLNIPYGINLPPLYRPNVTVP 60
QY 61 SYPGNTYTDGLSPYWLITSPGFYVYHIRGFPLATQNLNVPPLPRGFFVPPSRFFSA 120
Db 61 SYPGNTYTDGLSPYWLITSPGFYVYHIRGFPLATQNLNVPPLPRGFFVPPSRFFSA 120
QY 121 AAAPAAPPIAAEPAAAPLTATPVAAPAAAGAPVAAEPAAAPVGAEPAAAPVAAEPAA 180
Db 121 AAAPAAPPIAAEPAAAPLTATPVAAPAAAGAPVAAEPAAAPVGAEPAAAPVAAEPAA 180
QY 181 EAPVGVPEAAEPPSPAEPATAKPAAPPEPHSPSLEQANQ 219
Db 181 EAPVGVPEAAEPPSPAEPATAKPAAPPEPHSPSLEQANQ 219

RESULT 3
PCT-US02-08278-1040
; Sequence 1040, Application PC/TUS0208278
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc.
; TITLE OF INVENTION: Human Secreted Proteins
; FILE REFERENCE: PS902/PCT
; CURRENT APPLICATION NUMBER: PCT/US02/08278
; CURRENT FILING DATE: 2002-03-19
; PRIOR APPLICATION NUMBER: US 60/331,287
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/306,171
; PRIOR FILING DATE: 2001-07-19
; PRIOR APPLICATION NUMBER: US 60/277,340
```

```
; PRIOR FILING DATE: 2001-03-21
; NUMBER OF SEQ ID NOS: 1988
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1040
; LENGTH: 219
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US02-08278-1040

Query Match 100.0%; Score 1178; DB 1; Length 219;
Best Local Similarity 100.0%; Pred. No. 5.7e-76;
Matches 219; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLLWACIVCVAFARKRRFPFIGEDDNDGHLHPSLNIPYGINLPPLYRPNVTVP 60
Db 1 MKLLWACIVCVAFARKRRFPFIGEDDNDGHLHPSLNIPYGINLPPLYRPNVTVP 60
QY 61 SYPGNTYTDGLSPYWLITSPGFYVYHIRGFPLATQNLNVPPLPRGFFVPPSRFFSA 120
Db 61 SYPGNTYTDGLSPYWLITSPGFYVYHIRGFPLATQNLNVPPLPRGFFVPPSRFFSA 120
QY 121 AAAPAAPPIAAEPAAAPLTATPVAAPAAAGAPVAAEPAAAPVGAEPAAAPVAAEPAA 180
Db 121 AAAPAAPPIAAEPAAAPLTATPVAAPAAAGAPVAAEPAAAPVGAEPAAAPVAAEPAA 180
QY 181 EAPVGVPEAAEPPSPAEPATAKPAAPPEPHSPSLEQANQ 219
Db 181 EAPVGVPEAAEPPSPAEPATAKPAAPPEPHSPSLEQANQ 219

RESULT 4
US-09-922-469-2
; Sequence 2, Application US/09922469
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul A.
; TITLE OF INVENTION: SECRETED SALIVARY ZSTG63 POLYPEPTIDE
; FILE REFERENCE: 97-71
; CURRENT APPLICATION NUMBER: US/09/922,469
; CURRENT FILING DATE: 2001-08-03
; PRIOR APPLICATION NUMBER: US 60/124,820
; PRIOR FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 219
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-922-469-2

Query Match 100.0%; Score 1178; DB 23; Length 219;
Best Local Similarity 100.0%; Pred. No. 5.7e-76;
Matches 219; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLLWACIVCVAFARKRRFPFIGEDDNDGHLHPSLNIPYGINLPPLYRPNVTVP 60
Db 1 MKLLWACIVCVAFARKRRFPFIGEDDNDGHLHPSLNIPYGINLPPLYRPNVTVP 60
QY 61 SYPGNTYTDGLSPYWLITSPGFYVYHIRGFPLATQNLNVPPLPRGFFVPPSRFFSA 120
Db 61 SYPGNTYTDGLSPYWLITSPGFYVYHIRGFPLATQNLNVPPLPRGFFVPPSRFFSA 120
QY 121 AAAPAAPPIAAEPAAAPLTATPVAAPAAAGAPVAAEPAAAPVGAEPAAAPVAAEPAA 180
Db 121 AAAPAAPPIAAEPAAAPLTATPVAAPAAAGAPVAAEPAAAPVGAEPAAAPVAAEPAA 180
QY 181 EAPVGVPEAAEPPSPAEPATAKPAAPPEPHSPSLEQANQ 219
Db 181 EAPVGVPEAAEPPSPAEPATAKPAAPPEPHSPSLEQANQ 219

RESULT 5
US-09-922-480-2
```

```
; Sequence 2, Application US/09922480
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; APPLICANT: Adler, David A.
; TITLE OF INVENTION: SECRETED SALIVARY ZS1G63 POLYPEPTIDE
; FILE REFERENCE: 97-71
; CURRENT APPLICATION NUMBER: US/09/922,480
; PRIOR FILING DATE: 2001-08-03
; PRIOR APPLICATION NUMBER: US 60/124,820
; PRIOR FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 219
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-922-480-2

Query Match      100.0%; Score 1178; DB 23; Length 219;
Best Local Similarity 100.0%; Pred. No. 5.7e-76;
Matches 219; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLLWACIVCVAFARKRRFFIGEDDNDGHPHLSNIPYGINLPPLPYRPNVTVP 60
Db 1 MKLLWACIVCVAFARKRRFFIGEDDNDGHPHLSNIPYGINLPPLPYRPNVTVP 60

QY 61 SYPGNTYTDGLPSYPWILTSFGPPYVYHIGFPLATQLNVPLPPRGFPVPSRFFSA 120
Db 61 SYPGNTYTDGLPSYPWILTSFGPPYVYHIGFPLATQLNVPLPPRGFPVPSRFFSA 120

QY 121 AAAPAAPPIAAEPAAAPLTATPVAAEPAAGAPVAAEPAAPVGAEPAAEAPVAAEPAA 180
Db 121 AAAPAAPPIAAEPAAAPLTATPVAAEPAAGAPVAAEPAAPVGAEPAAEAPVAAEPAA 180

QY 181 EAPVGVEPAAEPSPAEPATAKPAAPPEHPSPSLEQANQ 219
Db 181 EAPVGVEPAAEPSPAEPATAKPAAPPEHPSPSLEQANQ 219

RESULT 6
US-09-923-236-2
; Sequence 2, Application US/09923236
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; APPLICANT: Adler, David A.
; TITLE OF INVENTION: SECRETED SALIVARY ZS1G63 POLYPEPTIDE
; FILE REFERENCE: 97-71
; CURRENT APPLICATION NUMBER: US/09/923,236
; CURRENT FILING DATE: 2001-08-03
; PRIOR APPLICATION NUMBER: US 60/124,820
; PRIOR FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 219
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-923-236-2

Query Match      100.0%; Score 1178; DB 23; Length 219;
Best Local Similarity 100.0%; Pred. No. 5.7e-76;
Matches 219; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLLWACIVCVAFARKRRFFIGEDDNDGHPHLSNIPYGINLPPLPYRPNVTVP 60
Db 1 MKLLWACIVCVAFARKRRFFIGEDDNDGHPHLSNIPYGINLPPLPYRPNVTVP 60

QY 61 SYPGNTYTDGLPSYPWILTSFGPPYVYHIGFPLATQLNVPLPPRGFPVPSRFFSA 120
Db 61 SYPGNTYTDGLPSYPWILTSFGPPYVYHIGFPLATQLNVPLPPRGFPVPSRFFSA 120

QY 121 AAAPAAPPIAAEPAAAPLTATPVAAEPAAGAPVAAEPAAPVGAEPAAEAPVAAEPAA 180
Db 121 AAAPAAPPIAAEPAAAPLTATPVAAEPAAGAPVAAEPAAPVGAEPAAEAPVAAEPAA 180

QY 181 EAPVGVEPAAEPSPAEPATAKPAAPPEHPSPSLEQANQ 219
Db 181 EAPVGVEPAAEPSPAEPATAKPAAPPEHPSPSLEQANQ 219

RESULT 7
US-10-009-328-29
; Sequence 29, Application US/10009328
; GENERAL INFORMATION:
; APPLICANT: INCYTE GENOMICS, INC.
; APPLICANT: IAL, Preeti
; APPLICANT: YANG, Junming
; APPLICANT: YUE, Henry
; APPLICANT: HILLMAN, Jennifer L.
; APPLICANT: TANG, Y. Tom
; APPLICANT: BANDMAN, Olga
; APPLICANT: BURFORD, Neil
; APPLICANT: BAUGHN, Mariah R.
; APPLICANT: AZIMZAI, Yalda
; APPLICANT: LU, Dyung Aina M.
; APPLICANT: AU-YOUNG, Janice
; APPLICANT: PATTERSON, Chandra
; TITLE OF INVENTION: HUMAN TRANSPORT PROTEINS
; FILE REFERENCE: PF-0709 PCT
; CURRENT APPLICATION NUMBER: US/10/009,328
; CURRENT FILING DATE: 2001-12-04
; PRIOR APPLICATION NUMBER: 60/139,923; 60/148,177; 60/149,357; 60/162,287
; PRIOR FILING DATE: 1999-06-17; 1999-08-10; 1999-08-18; 1999-10-28
; NUMBER OF SEQ ID NOS: 86
; SOFTWARE: PERL Program
; SEQ ID NO 29
; LENGTH: 219
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: misc.feature
; OTHER INFORMATION: Incyte ID No: 1325518CD1
US-10-009-328-29

Query Match      100.0%; Score 1178; DB 24; Length 219;
Best Local Similarity 100.0%; Pred. No. 5.7e-76;
Matches 219; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLLWACIVCVAFARKRRFFIGEDDNDGHPHLSNIPYGINLPPLPYRPNVTVP 60
Db 1 MKLLWACIVCVAFARKRRFFIGEDDNDGHPHLSNIPYGINLPPLPYRPNVTVP 60

QY 61 SYPGNTYTDGLPSYPWILTSFGPPYVYHIGFPLATQLNVPLPPRGFPVPSRFFSA 120
Db 61 SYPGNTYTDGLPSYPWILTSFGPPYVYHIGFPLATQLNVPLPPRGFPVPSRFFSA 120

QY 121 AAAPAAPPIAAEPAAAPLTATPVAAEPAAGAPVAAEPAAPVGAEPAAEAPVAAEPAA 180
Db 121 AAAPAAPPIAAEPAAAPLTATPVAAEPAAGAPVAAEPAAPVGAEPAAEAPVAAEPAA 180

QY 181 EAPVGVEPAAEPSPAEPATAKPAAPPEHPSPSLEQANQ 219
Db 181 EAPVGVEPAAEPSPAEPATAKPAAPPEHPSPSLEQANQ 219

RESULT 8
US-60-162-287-4
; Sequence 4, Application US/60162287
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Tang, Y. Tom
; APPLICANT: Burford, Neil
; APPLICANT: Yue, Henry
; APPLICANT: Lal, Preeti
; APPLICANT: Patterson, Chandra
; APPLICANT: Baughn, Mariah R.
```

```

; APPLICANT: Lu, Dyung Aina M.
; TITLE OF INVENTION: TRANSPORT PROTEINS
; FILE REFERENCE: PF-0748 P
; CURRENT APPLICATION NUMBER: US/60/162,287
; CURRENT FILING DATE: 1999-10-28
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: PERL Program
; SEQ ID NO 4
; LENGTH: 219
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No: 1325518CD1
; US-60-162-287-4

```

	Query Match	100.0%	Score 1178;	DB 27;	Length 219;
	Best Local Similarity	100.0%;	Pred. No. 5.7e-76;		
	Matches 219;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	MKLLWACIVCAVAFARKRRFFPIGEDDNDGCHLPHSLNIPYGI RNLPPLXYRYPVTVP	60		
Dd	1	MKLLLWACIVCAVAFARKRRFFPIGEDDNDGCHLPHSLNIPYGI RNLPPLXYRYPVTVP	60		
Qy	61	SYPGNTYTTDTGLPSYPWLITSPGFVVYHIRGFPLATQLNV PPLPRGPFPPVPSRFFSA	120		
Dd	61	SYPGNTYTTDTGLPSYPWLITSPGFVVYHIRGFPLATQLNV PPLPRGPFPPVPSRFFSA	120		
Qy	121	AAPAAPPTAAEPAAAAA PLTATPVAAEPAACAGPVAAEPA AFAVPVGAEPAEAAPVAEAPAA	180		
Dd	121	AAPAAPPTAAEPAAAAA PLTATPVAAEPAACAGPVAAEPA AFAVPVGAEPAEAAPVAEAPAA	180		
Qy	181	EAPVGVEPAEEPSPAEPATAKPAAPEHPSPSLEFAQO	219		
Dd	181	FAPGVGEPAERPSPAEPATAKPAAPEHPSPSLEFAQO	219		

```

RESULT 9
PCT-US98-13813-10
: Sequence 10, Application PC/TUS9813813
: GENERAL INFORMATION:
: APPLICANT: Jacobs, Kenneth
: APPLICANT: McCoy, John M.
: APPLICANT: LaVallie, Edward R.
: APPLICANT: Racie, Lisa A.
: APPLICANT: Treacy, Maurice
: APPLICANT: Spaulding, Vikki
: APPLICANT: Agostino, Michael J.
: TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
: TITLE OF INVENTION: ENCODING THEM
: NUMBER OF SEQUENCES: 32
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Genetics Institute, Inc.
: STREET: 87 CambridgePark Drive
: CITY: Cambridge
: STATE: MA
: COUNTRY: U.S.A.
: ZIP: 02140
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US98/13813
: FILING DATE:
: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
: NAME: Sprunger, Suzanne A.
: REGISTRATION NUMBER: 41,323
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (617) 498-8284
: TELEFAX: (617) 876-5851

```

```
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 219 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCIT-US98-L3813-10

Query Match          99.5%; Score 1172; DB 1; Length 219;
Best Local Similarity 99.1%; Pred. No. 1.5e-75;
Matches 217; Conservative 2; Mismatches 0; Indels 0; Gaps

QY   1 MKLLLNACIVCVAFARRRPFPTIGEDDNDGPHLHPSLNIPYGIRNLPPLYXRPVNTVP 60
      |||
Db    1 MKLLLNACIVCVAFARRRPFPTIGEDDNDGPHLHPSLNIPYGIRNLPPLYXRPVNTVP 60

QY   61 SYPGTNTDTGLSPWILTSFGFVVYHIRGFPLATQLNVPLPRGRGFFVPPSRFFSA 120
      |||
Db    61 SYPGTNTDTGLSPWILTSFGFVVYHIRGFPLATQLNVPLPRGRGFFVPPSRFFSA 120

QY   121 AAAPAAPPIAAEPAAAAPLTATVAEPAAGAPVAAEPAEAEPVGAEPAEAEPVAAEPA 180
      |||
Db    121 AAAPAAPPIAAEPAAAAPLTSTVASEPAAGAPVAAEPAEAEPVGAEPAEAEPVAAEPA 180

QY   181 EAPGVGEPAEEPSAEPATAKPAAPEHPSPSLEQANQ 219
      |||
Db    181 EAPGVGEPAEEPSAEPATAKPAAPEHPSPSLEQANQ 219

RESULT 10
US-08-958-304-10
; Sequence 10, Application US/08958304
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John M.
; APPLICANT: Lavallie, Edward R.
; APPLICANT: Racine, Lisa A.
; APPLICANT: Treacy, Maurice
; APPLICANT: Spaulding, Vikki
; APPLICANT: Agostino, Michael J.
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
; TITLE OF INVENTION: ENCODING THEM
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
```

ZIP: 02140
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent In Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/958,304
 FILING DATE:
 CLASSIFICATION: 536
 ATTORNEY/AGENT INFORMATION:
 NAME: Sprunger, Suzanne A.
 REGISTRATION NUMBER: 41,323
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (617) 498-8284
 TELEFAX: (617) 876-5851
 INFORMATION FOR SEQ ID NO: 10:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 219 amino acids
 TYPE: amino acid
 STRANDEDNESS:
 TOPOLOGY: linear
 MOLECULE TYPE: protein

50:00

Query Match	99.5%	Score 1172;	DB 21;	Length 219;
Best Local Similarity	99.1%	Pred. No. 1.5e-75;		

PCT-US02-08277-1086
; Sequence 1086, Application PC/TUS0208277
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc.
; TITLE OF INVENTION: Human Secreted Protein

```

; FILE REFERENCE: PS907PCT
; CURRENT APPLICATION NUMBER: PCT/US02/08277
; CURRENT FILING DATE: 2002-03-19
; PRIOR APPLICATION NUMBER: US 60/331,287
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/306,171
; PRIOR FILING DATE: 2001-07-19
; PRIOR APPLICATION NUMBER: US 60/277,340
; PRIOR FILING DATE: 2001-03-21
; NUMBER OF SEQ ID NOS: 1357
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1086
; LENGTH: 220
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (159)
; OTHER INFORMATION: Xaa equals any amino acid
PCT-US02-08277-1086

Query Match          94.5%; Score 1113.5; DB 1; Length 220;
Best Local Similarity 95.0%; Pred. No. 2.3e-71;
Matches 209; Conservative 1; Mismatches 9; Indels 1; Gaps 1;

QY 1 MKLLWACIVCVAFARKRREFFIGEDDDGHLHPSLNIPYGI RNLPPPLYRPNVTVP 60
DB 1 MKLLWACIVCVAFARKRREFFIGEDDDGHLHPSLNIPYGI RNLPPPLYRPNVTVP 60

QY 61 SYPGNTYTDGLPSYWI L TSPGFVYVYHIRGFP L ATQ L NVPLP P R G F F V P P S R F F S A 120
DB 61 SYPGNTYTDGLPSYWI L TSPGFVYVYHIRGFP L ATQ L NVPLP P R G F F V P P S R F F S A 120

QY 121 AAAPAAPPIAAEPA A A A A P L T A T P V A A E P A A G A P V A A E P A A E A P - V G A E P A A E A P V A A E P A 179
DB 121 AAAPAAPPIAAEPA A A A A P L T A T P V A A E P A A G A P V A A E P A A E A P - V G A E P A A E A P V A A E P A 180

QY 180 AEAPVGVEPA A E E P S P A E P A T A K P A P E P H P S P S L E Q A N Q 219
DB 181 AEAPVGVEPA A E E P S P A E P A T A K P A P E P H P S P S L E Q A N Q 220

RESULT 15
PCT-US98-22376-190
; Sequence 190, Application PC/TUS9822376
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc.
; TITLE OF INVENTION: 148 Human Secreted Proteins
; FILE REFERENCE: P2019.PCT
; CURRENT APPLICATION NUMBER: PCT/US98/22376
; CURRENT FILING DATE: 1998-10-23
; EARLIER APPLICATION NUMBER: 60/063,099
; EARLIER FILING DATE: 1997-10-24
; EARLIER APPLICATION NUMBER: 60/063,088
; EARLIER FILING DATE: 1997-10-24
; EARLIER APPLICATION NUMBER: 60/063,387
; EARLIER FILING DATE: 1997-10-24
; EARLIER APPLICATION NUMBER: 60/063,148
; EARLIER FILING DATE: 1997-10-24
; EARLIER APPLICATION NUMBER: 60/063,386
; EARLIER FILING DATE: 1997-10-24
; EARLIER APPLICATION NUMBER: 60/062,784
; EARLIER FILING DATE: 1997-10-24
; EARLIER APPLICATION NUMBER: 60/063,091
; EARLIER FILING DATE: 1997-10-24
; EARLIER APPLICATION NUMBER: 60/063,090
; EARLIER FILING DATE: 1997-10-24
; EARLIER APPLICATION NUMBER: 60/063,089
; EARLIER FILING DATE: 1997-10-24
; EARLIER APPLICATION NUMBER: 60/063,092
; EARLIER FILING DATE: 1997-10-24
; EARLIER APPLICATION NUMBER: 60/063,109
; EARLIER FILING DATE: 1997-10-24
; EARLIER APPLICATION NUMBER: 60/063,111
; EARLIER FILING DATE: 1997-10-24
; EARLIER APPLICATION NUMBER: 60/063,101
; EARLIER FILING DATE: 1997-10-24
; EARLIER APPLICATION NUMBER: 60/063,109
; EARLIER FILING DATE: 1997-10-24
; EARLIER APPLICATION NUMBER: 60/063,110
; EARLIER FILING DATE: 1997-10-24
; EARLIER APPLICATION NUMBER: 60/063,097
; EARLIER FILING DATE: 1997-10-24
; EARLIER APPLICATION NUMBER: 60/063,100
; EARLIER FILING DATE: 1997-10-24
; EARLIER APPLICATION NUMBER: 60/063,098
; EARLIER FILING DATE: 1997-10-24
; NUMBER OF SEQ ID NOS: 492
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 190
; LENGTH: 221
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (159)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (221)
;
Query Match          94.5%; Score 1113.5; DB 1; Length 220;
Best Local Similarity 95.0%; Pred. No. 2.3e-71;
Matches 209; Conservative 1; Mismatches 9; Indels 1; Gaps 1;

PCT-US02-08278-1563
; Sequence 1563, Application PC/TUS0208278
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc.
; TITLE OF INVENTION: Human Secreted Proteins
; FILE REFERENCE: PS902PCT
; CURRENT APPLICATION NUMBER: PCT/US02/08278
; CURRENT FILING DATE: 2002-03-19
; PRIOR APPLICATION NUMBER: US 60/331,287
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/306,171
; PRIOR FILING DATE: 2001-07-19
; PRIOR APPLICATION NUMBER: US 60/277,340
; PRIOR FILING DATE: 2001-03-21
; NUMBER OF SEQ ID NOS: 1988
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1563
; LENGTH: 220
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (159)
; OTHER INFORMATION: Xaa equals any amino acid
PCT-US02-08278-1563

Query Match          94.5%; Score 1113.5; DB 1; Length 220;
Best Local Similarity 95.0%; Pred. No. 2.3e-71;
Matches 209; Conservative 1; Mismatches 9; Indels 1; Gaps 1;
```

```
; OTHER INFORMATION: Xaa equals stop translation
PCT-US98-22376-190
Query Match          94.5%; Score 1113.5; DB 1; Length 221;
Best Local Similarity 95.0%; Pred. No. 2.3e-71;
Matches 209; Conservative 1; Mismatches 9; Indels 1; Gaps 1;

QY 1 MKLLWACIVCVAFARKRRFFFIGEDDNDGHLHPSLNIPYGI RNLPPLYYRPVNTVP 60
   ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1 MKLLWACIVCVAFARKRRFFFIGEDDNDGHLHPSLNIPYGI RNLPPLYYRPVNTVP 60
   ||||||||||||||||||||||||||||||||||||||||||||||||||||

QY 61 SYPGNTYTDGLPSYPWILTSFGFPYVYHIRGEPLATQLNVPPLPRGFPVPSPSREFSA 120
   ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 61 SYPGNTYTDGLPSYPWILTSFGFPYVYHIRGEPLATQLNVPPLPRGFPVPSPSREFSA 120
   ||||||||||||||||||||||||||||||||||||||||||||||||||||

QY 121 AAAPAAPPIAAEPAAAPLTATPVAAEPAAAGAPVAAEPAAEAP-VGAEPAAEAPVAAEPA 179
   ||||||||||||||||||||||||||||||||||||||||||||||||: ||||||||||||
Db 121 AAAPAAPPIAAEPAAAPLTATPVAAEPAAAGAPVAAEPAAEAPVGAEPAAEAPVAAEPA 180
   ||||||||||||||||||||||||||||||||||||||||||||||||

QY 180 AEAPVGVPEAAEPEPSAPATAKPAAPPEPHPSPSLEQANQ 219
   ||||||||||||||||||||||||||||||||||||||||||||||||
Db 181 AEAPVGVPEAAEPEPSAPATAKPAAPPEPHPSPSLEQANQ 220
   ||||||||||||||||||||||||||||||||||||||||||||||||
```

Search completed: February 12, 2003, 13:53:45
Job time : 245 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2003, 13:46:29 ; Search time 95 Seconds
(without alignments)
193.152 Million cell updates/sec

Title: US-09-923-236-2
Perfect score: 1178
Sequence: 1 MKLLWACIVCAFARKRF.....TAKPAPEPHPSLEQANQ 219

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 425801 seqs, 83787487 residues

Total number of hits satisfying chosen parameters: 425801

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Pending_Patents_AA_New.*

- 1: /cgn2_6/ptodata/1/paa/PCT_NEW_COMB.pep.*
- 2: /cgn2_6/ptodata/1/paa/US06_NEW_COMB.pep.*
- 3: /cgn2_6/ptodata/1/paa/US07_NEW_COMB.pep.*
- 4: /cgn2_6/ptodata/1/paa/US08_NEW_COMB.pep.*
- 5: /cgn2_6/ptodata/1/paa/US09_NEW_COMB.pep.*
- 6: /cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep.*
- 7: /cgn2_6/ptodata/1/paa/US60_NEW_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Match	Length	DB	ID	Description
1	181	15.4	2657	6	US-10-316-253-95		Sequence 95, Appl
2	175.5	14.9	915	5	US-09-724-676A-72183		Sequence 72183, A
3	175.5	14.9	915	5	US-09-724-676A-72183		Sequence 72183, A
4	170	14.4	428	6	US-10-198-723-14		Sequence 14, Appl
5	170	14.4	1028	6	US-10-099-056-1091		Sequence 1091, Ap
6	156	13.2	2134	5	US-09-724-676A-61374		Sequence 61374, A
7	156	13.2	2134	5	US-09-724-676A-61374		Sequence 61374, A
8	156	13.2	5159	6	US-10-085-198-112		Sequence 112, App
9	155	13.2	993	1	PCT-US02-39133-7		Sequence 7, Appl
10	148	12.6	2971	6	US-10-218-140-1990		Sequence 1990, Ap
11	147	12.5	497	6	US-10-348-119-337		Sequence 337, App
12	145.5	12.4	441	6	US-10-276-774-1783		Sequence 1783, Ap
13	145.5	12.4	1458	6	US-10-312-352-27		Sequence 27, Appl
14	145.5	12.4	1489	1	PCT-US02-38437-13		Sequence 13, Appl
15	144.5	12.3	271	5	US-09-724-676A-89247		Sequence 89247, A
16	144.5	12.3	271	5	US-09-724-676A-89247		Sequence 89247, A
17	144.5	12.3	271	5	US-09-724-676A-89249		Sequence 89249, A
18	144.5	12.3	271	5	US-09-724-676A-89249		Sequence 89249, A
19	144.5	12.3	271	5	US-09-724-676A-89250		Sequence 89250, A
20	144.5	12.3	271	5	US-09-724-676A-89247		Sequence 89247, A
21	144.5	12.3	271	5	US-09-724-676A-89248		Sequence 89248, A
22	144.5	12.3	271	5	US-09-724-676A-89249		Sequence 89249, A
23	144.5	12.3	281	5	US-09-724-676A-89250		Sequence 89250, A
24	144.5	12.3	281	5	US-09-724-676A-89223		Sequence 89223, A
25	144.5	12.3	281	5	US-09-724-676A-89224		Sequence 89224, A
26	144.5	12.3	281	5	US-09-724-676A-89225		Sequence 89225, A
					US-09-724-676-89226		Sequence 89226, A

27	144.5	12.3	281	5	US-09-724-676A-89223		Sequence 89223, A
28	144.5	12.3	281	5	US-09-724-676A-89224		Sequence 89224, A
29	144.5	12.3	281	5	US-09-724-676A-89225		Sequence 89225, A
30	144.5	12.3	281	5	US-09-724-676A-89226		Sequence 89226, A
31	144.5	12.3	452	5	US-09-724-676-93540		Sequence 93540, A
32	144.5	12.3	452	5	US-09-724-676-93545		Sequence 93545, A
33	144.5	12.3	452	5	US-09-724-676A-93540		Sequence 93540, A
34	144.5	12.3	452	5	US-09-724-676A-93545		Sequence 93545, A
35	144.5	12.3	632	5	US-09-724-676-89194		Sequence 89194, A
36	144.5	12.3	632	5	US-09-724-676-89195		Sequence 89195, A
37	144.5	12.3	632	5	US-09-724-676-89196		Sequence 89196, A
38	144.5	12.3	632	5	US-09-724-676-89197		Sequence 89197, A
39	144.5	12.3	632	5	US-09-724-676A-89194		Sequence 89194, A
40	144.5	12.3	632	5	US-09-724-676A-89195		Sequence 89195, A
41	144.5	12.3	632	5	US-09-724-676A-89196		Sequence 89196, A
42	144.5	12.3	632	5	US-09-724-676A-89197		Sequence 89197, A
43	144.5	12.3	639	5	US-09-724-676-89175		Sequence 89175, A
44	144.5	12.3	639	5	US-09-724-676-89176		Sequence 89176, A
45	144.5	12.3	639	5	US-09-724-676A-89175		Sequence 89175, A

ALIGNMENTS

RESULT 1
US-10-316-253-95
; Sequence 95, Application US/10316253
; GENERAL INFORMATION:
; APPLICANT: The Procter & Gamble Company
; APPLICANT: Peters, Kevin
; APPLICANT: Thompson, Larry
; APPLICANT: Wang, Feng
; APPLICANT: Greis, Kenneth
; TITLE OF INVENTION: Angiogenesis Modulating Proteins
; FILE REFERENCE: 8865M
; CURRENT APPLICATION NUMBER: US/10/316,253
; PRIOR FILING DATE: 2002-12-10
; PRIOR APPLICATION NUMBER: US 60/355,295
; NUMBER OF SEQ ID NOS: 308
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 95
; LENGTH: 2657
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-316-253-95

Query Match 15.4%; Score 181; DB 6; Length 2657;
Best Local Similarity 29.1%; Pred. No. 0.0051;
Matches 57; Conservative 22; Mismatches 61; Indels 56; Gaps 8;
QY 40 IPYGIRNLPPP-----LYRPNVTVPSPGNTYDTGLPSYPMILTSPPGPPVYHIRGF 93
Db 2239 VKFGHQINPHTFANESLTSKVVTIK--PSNSSKPTAI-----VNLTP----- 2280
QY 94 PLATQNLVPLPPRGRPFVFP-----PSRFEFAAAAPAP-----PIAAEP 133
Db 2281 -----AKPAPAPAPQAPVLAQPKAPQAPAPAKLPPQPPVHVQAPAQ 2333
QY 134 AAAAPLTPVAAEPAAGAPVAAEPAEAAEA-PVGAEPAAEAPVAAEPAEAPVGVPEAAEE 192
Db 2334 ASVRPAPAPAPQAPAAKPVPAKPAVPAQAPAPQAPAAKPVPAKPAVPA----QPAAP 2389
QY 193 PSAPAEATAPAPAP 208
Db 2390 PMPAQPVLTAKVP 2405

RESULT 2
US-09-724-676-72183
; Sequence 72183, Application US/09724676
; GENERAL INFORMATION:
; APPLICANT: CompuGen LTD


```

; APPLICANT: GRIFFIN, Jennifer A
; APPLICANT: RAMKUMAR, Jayalaxmi
; APPLICANT: EMERLING, Brooke M.
; APPLICANT: KABLE, Amy E.
; APPLICANT: ELLIOTT, Vicki S.
; APPLICANT: MARQUIS, Joseph P.
; APPLICANT: BAUGHN, Mariah R.
; APPLICANT: GORVAD, Ann E.
; APPLICANT: YUE, Henry
; APPLICANT: LEE, Ernestine A.
; APPLICANT: BECHA, Shanya D.
; APPLICANT: TANG, Y. Tom
; APPLICANT: TRAN, Uyen K.
; APPLICANT: SWARNAKAR, Anita
; APPLICANT: LEE, Sally
; APPLICANT: ISON, Craig H.
; APPLICANT: HAFALIA, April J.A.
; APPLICANT: TRAN, Bao
; APPLICANT: SPRAGUE, William W.
; APPLICANT: LEE, Soo Yeun
; APPLICANT: GANDHI, Aneena R.
; APPLICANT: GIETZEN, Kimberly J.
; APPLICANT: BHATIA, Umesh G.
; APPLICANT: BURRILL, John D.
; APPLICANT: BLAKE, Julie J.
; APPLICANT: HO, Anne
; APPLICANT: Zheng, Wenjin
; TITLE OF INVENTION: PROTEINS ASSOCIATED WITH CELL GROWTH, DIFFERENTIATION, AND DEATH
; FILE REFERENCE: PF-1319 PCT
; CURRENT APPLICATION NUMBER: PCT/US02/39133
; CURRENT FILING DATE: 2002-12-04
; PRIOR APPLICATION NUMBER: US 60/340,747
; PRIOR FILING DATE: 2001-12-07
; PRIOR APPLICATION NUMBER: US 60/342,761
; PRIOR FILING DATE: 2001-12-20
; PRIOR APPLICATION NUMBER: US 60/349,705
; PRIOR FILING DATE: 2002-01-15
; PRIOR APPLICATION NUMBER: US 60/354,764
; PRIOR FILING DATE: 2002-02-06
; PRIOR APPLICATION NUMBER: US 60/356,216
; PRIOR FILING DATE: 2002-02-12
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PERL Program
; SEQ ID NO 7
; LENGTH: 993
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No: 5137816CD1
PCT-US02-39133-7

Query Match      13.2%; Score 155; DB 1; Length 993;
Best Local Similarity 40.2%; Pred. No. 0.073;
Matches 47; Conservative 6; Mismatches 44; Indels 20; Gaps 4;

QY 121 AAAPAPPIAEPAAAG---PLT--ATPVAEPAAAG--PVAEPAAEPAAEPAAEPAAEP 173
Db 818 AAAPVAAAAAPAAAAAPSPATAATAANVSPAAGQIPAAASVASAAAVAPSAAAAA 877

QY 174 VAAEPAAEPVGEPAEEPSPAEPATAKPA-----APEPHSPSPLEQA 217
Db 878 VQVAPAAPAPVPAPALVPVPAPAPAAQAASAPAQTAQTSAPAVAPTPATPTTPAQA 934

RESULT 10
US-10-218-140-1990
; Sequence 1990, Application US/10218140
; GENERAL INFORMATION:
; APPLICANT: Leach, Martin D.
; APPLICANT: Shmukets, Richard A.
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES AND POLYPEPTIDES

```

```

; TITLE OF INVENTION: ENCODED THEREBY
; FILE REFERENCE: 15966-543 CON
; CURRENT APPLICATION NUMBER: US/10/218,140
; CURRENT FILING DATE: 2002-08-12
; PRIOR APPLICATION NUMBER: 09/540,763
; PRIOR FILING DATE: 2000-03-30
; PRIOR APPLICATION NUMBER: 60/127,728
; PRIOR FILING DATE: 1999-04-05
; PRIOR APPLICATION NUMBER: 60/127,636
; PRIOR FILING DATE: 1999-04-02
; PRIOR APPLICATION NUMBER: 60/127,607
; PRIOR FILING DATE: 1999-03-31
; NUMBER OF SEQ ID NOS: 6322
; SOFTWARE: Curanator Version 1.0
; SEQ ID NO 1990
; LENGTH: 2971
; TYPE: PRT
; ORGANISM: Homo sapiens
; ORGANISM: Homo sapiens
US-10-218-140-1990

Query Match      12.6%; Score 148; DB 6; Length 2971;
Best Local Similarity 26.9%; Pred. No. 0.48;
Matches 63; Conservative 24; Mismatches 91; Indels 56; Gaps 10;

QY 33 PLHPSLNIPYGI RNLPPLPYRVNTVPSYPG-----NTYTDGLPSYPMILT 80
Db 1203 PLTVSASGPALLTSVTPL--APV--VPAAPGPPSLAPSGASPSASALTGLATAPSLSS 1258

QY 81 S--PGFPYVY-----HTRGF-----PLATOLNVP-----PLPRGFPVPVPS 115
Db 1259 SQTGPHPLLLAPTSSHVPLNSTVAPACSEVLVPASALASPFPSAPNPAQAASLLAPAS 1318

QY 116 RFESAAPAAAPPIAAB-----PAAAPLTATPVAAEPAAAGAPVAAEPAAEPV----- 164
Db 1319 SASQALATPLAPMAAPQTAILAFSPAPPLAPLPVLAPSGAAPVLASSQTPVPVMAPSST 1378

QY 165 -GAEPAAEPVAAEPAAEPVGEPAEEPSA---EPATAKPAEPPEHPSPSL 214
Db 1379 PGTSLASASPVAPTVPVLAPSSQTOTMLPADVPSPPLSPASTQTALALAPALAPT 1432

RESULT 11
US-10-348-119-337
; Sequence 337, Application US/10348119
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: IDENTIFICATION OF POLYNUCLEOTIDES AND POLYPEPTIDE FOR PREDICTI
; TITLE OF INVENTION: ACTIVITY OF COMPOUNDS THAT INTERACT WITH PROTEIN TYROSINE KIN
; TITLE OF INVENTION: AND/OR PROTEIN TYROSINE KINASE PATHWAYS
; FILE REFERENCE: D0185 NP
; CURRENT APPLICATION NUMBER: US/10/348,119
; CURRENT FILING DATE: 2003-01-17
; PRIOR APPLICATION NUMBER: US 60/350,061
; PRIOR FILING DATE: 2002-01-18
; NUMBER OF SEQ ID NOS: 795
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 337
; LENGTH: 497
; TYPE: PRT
; ORGANISM: Homo sapiens
; ORGANISM: Homo sapiens
US-10-348-119-337

Query Match      12.5%; Score 147; DB 6; Length 497;
Best Local Similarity 32.8%; Pred. No. 0.12;
Matches 62; Conservative 13; Mismatches 82; Indels 32; Gaps 8;

QY 54 RPVNTVPSYP-GNTYTDGLPSYPMILTSPGFYVYHIRGFPLATOLNVPPPLPRGFPFV 112
Db 9 RPAAPRPRWSLGPACAEAMFVPEPCTLLAPPFP-VLGLDSRGVGLMNSFP-PPQGHQAO 66

QY 113 P-----PSRFFSA-----AAAPAAAPPI-----AAEP-----AAAAPLTATPV 144
Db 67 PLQVGAELQSRFFASQCGCAQSPFQAAPAPPTTQAPAAEPQLQVDLLPLVLAQAQESAAAA 126

```

[illegible]

```

RESULT 12
US-10-276-774-1783
; Sequence 1783, Application US/10276774
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; APPLICANT: Tang, Y, Tom et al
; TITLE OF INVENTION: Novel Nucleic Acids and Polypeptides
; FILE REFERENCE: 21272-030
; CURRENT APPLICATION NUMBER: US/10/276,774
; CURRENT FILING DATE: 2002-11-18
; PRIOR APPLICATION NUMBER: 09/560,875
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 09/496,914
; PRIOR FILING DATE: 2000-02-03
; NUMBER OF SEQ ID NOS: 2700
; SOFTWARE: Custom
; SEQ ID NO 1783
; LENGTH: 441
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-276-774-1783

```

```

Query Match      12.4%; Score 145.5; DB 6; Length 441;
Best Local Similarity 28.9%; Pred. No. 0.13;
Matches 61; Conservative 14; Mismatches 95; Indels 41; Gaps 7;

QY    47 LPPLLYRVNTVPSPNGT-----YTDGLPSYPMLTSPGFPVYHIRGFPLAT 97
      ||| || : :: || : || | | | | | | | | | | | | | | | | |
Db     21 LPPLLQQLLHSLGAPGESILLVRTSKLLVLGCLQLLVLLQTLSLLALQHITSSAP 80

QY    98 QLNVP-----PLPRGPFVPPSRFFSAAA--APAAPTAAEPAAAA 136
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     81 LLAAPTAVGSCSRCAPSRSRCVARPAAARTGLPTAPASSPAPASPAPASPAEASTAP 140

QY    137 APLTATPVAAEPAAGAP--VAAGPAAEAPVGAEPAEAAPVAAEPAAEPVGEPAEEP 193
       || | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     141 QPILLLP-KPPAPAGAPPRPGCPPRPAASPSPAAPPAAASPVLTASPPPLPAASPSP 199

QY    194 --SPAEP-----TAKPAAEPHPFSLEQA 217
       || | | | | | | | | | | | | | | | | | | | | | | | |
Db     200 AASPAPPAASPVLTASPPPLPAASPSPASPA 230

```

```

RESULT 13
US-10-312-352-27
Sequence 27, Application US/10312352
GENERAL INFORMATION:
APPLICANT: INCYTE GENOMICS, INC.; TANG, Y. TOM
APPLICANT: YUE, Henry; AZIMZAI, Yalda
APPLICANT: HE, Ann; BAIRA, Sajeev
APPLICANT: HO, Terence P.; NGUYEN, Danniell B.
APPLICANT: BURRILL, John D.; MARCUS, Gregory A.
APPLICANT: ZINGLER, Kurt A.; GANDHI, Ameen A.
APPLICANT: LAL, Preeti G.; KEARNEY, Liam
APPLICANT: BURFORD, Neil; YAO, Monique G.
APPLICANT: CHAMLA, Narinder K.; ELLIOT, Vicki S.
APPLICANT: ARVIZU, Chandra S.; KHAN, Farrah A.
APPLICANT: BAUGHN, Mariah R.; HAFALIA, April, J.A.
APPLICANT: POLICKY, Jennifer L.; AU-YOUNG, Janice K.
APPLICANT: LU, Yan; BOROWSKY, Mark L.
APPLICANT: LU, Dyung Aina M.; RAMKOMAR, Jayalaxmi
APPLICANT: YANG, Junming; GURURAJAN, Rajagopal
APPLICANT: WARREN, Bridget A.; GIETZEN, Kimberly J.

```

```

: APPLICANT: XU, Yuming; KALLICK, Deborah A.
: APPLICANT: LEE, Ernestine A.; THANGAVELU, Kavitha
: APPLICANT: DELEGEANE, Angelo M.; LEE, Sally
: TITLE OF INVENTION: EXTRACELLULAR MATRIX AND CELL ADHESION MOLECULES
: FILE REFERENCE: PF-0794 USN
: CURRENT APPLICATION NUMBER: US/10/312,352
: CURRENT FILING DATE: 2002-12-18
: PRIOR APPLICATION NUMBER: PCT/US01/21067
: PRIOR FILING DATE: 2001-06-29
: PRIOR APPLICATION NUMBER: US 60/215,454
: PRIOR FILING DATE: 2000-06-30
: PRIOR APPLICATION NUMBER: US 60/219,462
: PRIOR FILING DATE: 2000-07-18
: PRIOR APPLICATION NUMBER: US 60/240,111
: PRIOR FILING DATE: 2000-10-12
: PRIOR APPLICATION NUMBER: US 60/240,106
: PRIOR FILING DATE: 2000-10-12
: PRIOR APPLICATION NUMBER: US 60/244,021
: PRIOR FILING DATE: 2000-10-27
: PRIOR APPLICATION NUMBER: US 60/248,887
: PRIOR FILING DATE: 2000-11-14
: PRIOR APPLICATION NUMBER: US 60/249,570
: PRIOR FILING DATE: 2000-11-16
: NUMBER OF SEQ ID NOS: 72
: SOFTWARE: PERL Program
: SEQ ID NO 27
: LENGTH: 1458
: TYPE: PRT
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: misc_feature
: OTHER INFORMATION: Incyte ID No: 4755911CD1
US-10-312-352-27

Query Match      12.4%; Score 145.5; DB 6; Length 1458;
Best Local Similarity 34.5%; Pred. No. 0.37;
Matches 61; Conservative 10; Mismatches 83; Indels 23; Gaps

QY      36  PSLNIPVGINLP--PPLYRYPVNT--VPSYPGN-TYDITGLPSYPWILTSGFPVYVHI 90
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      1236  PPLPPPEGEARKAKAPVPSGKPVLTQVPKLOGSTPTTSKKVP-----LPDGGSPFVKRA 1290

QY      91  RGFLPLATQLNVVPLP--PRGFPVFPSPRFSAAAAAPAAAPPAEPAAPLTAATPVAEEP 148
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      1291  HGTTPPVPSKPPPPPTAPKPVKAVAGLPGSSAGSPAPSPARQPPAALAKPPTP----P 1346

QY      149  AAGAPVAAPAEAP---VGAPEAAEPVAAPEAAEPVGVPEAAEESPAPETAK 202
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      1347  SLGASAKPESPGALHPYAKPPPPRAAAAAAAAAPPAPEGA----SPGDSAROK 1399

```

```

RESULT 14
PCT-US02-38437-13
; Sequence 13, Application PC/TUS0238437
; GENERAL INFORMATION:
; APPLICANT: INCYTE GENOMICS, INC.
; APPLICANT: BAUGHN, Mariah R.
; APPLICANT: BECHTA, Shanya D.
; APPLICANT: BHATIA, Umesh G.
; APPLICANT: BLAKE, Julie J.
; APPLICANT: BOROWSKY, Mark L.
; APPLICANT: BURRILL, John D.
; APPLICANT: DELEGEANE, Angelo M.
; APPLICANT: ELLIOTT, Vicki S.
; APPLICANT: GANDHI, Ameena R.
; APPLICANT: GIETZEN, Kimberly J.
; APPLICANT: GORVAD, Ann E.
; APPLICANT: GRIFFIN, Jennifer A.
; APPLICANT: HO, Anne
; APPLICANT: JIN, Pei
; APPLICANT: KABLE, Amy E.
; APPLICANT: IAL, Preeti G.
; APPLICANT: LEE, Ernestine A.

```


Result No.	Query No.	Score	%		Length	DB	ID	Description
			Match					
1	206.5	17.5	607	2	S27776	80K protein (allele		
2	199	16.9	391	2	S27850	surface antigen Tc		
3	193	16.4	3534	2	T42567	tegument protein 2		
4	188.5	16.0	1108	2	T18353	protein p97 - Myc		
5	188	16.0	839	2	F75518	hypothetical prote		
6	186.5	15.8	395	2	H75457	hypothetical prote		
7	183	15.5	744	2	E86355	hypothetical prote		
8	179	15.2	207	2	A40215	hypothetical prote		
9	175	14.9	1054	2	T43226	TcB antigen - Tryp		
10	175	14.9	3149	1	Q08BE8	translation initia		
11	174.5	14.8	1611	2	T38236	BPLF1 protein - hu		
12	173.5	14.7	786	2	T01456	hypothetical prote		
13	172.5	14.6	846	2	S52418	extensin homolog F		
14	171.5	14.6	438	2	G87675	GRP-binding regul		
15	171	14.5	544	2	T17547	arylesterase-relat		
16	170.5	14.5	601	2	S27777	proline-rich prote		
17	170	14.4	548	2	E70546	80K protein (alle		
18	170	14.4	873	2	A47283	hypothetical prote		
19	170	14.4	3176	1	CGH03A	calphostin - fruit		
20	169	14.3	865	2	A47282	collagen alpha 3(V		
21	169	14.3	1099	2	T18357	calcium-binding pr		
22	168.5	14.3	217	2	S29309	mhlpl protein - Myc		
23	168.5	14.3	309	2	G83013	hypothetical prote		
24	167.5	14.2	450	1	C29413	polyhydroxyalkano		
25	165.5	14.0	416	1	SKXLAG	ubiquinol-cytochr		
26	165.5	14.0	416	2	T34279	dermal gland prote		
27	165.5	14.0	801	2	T29018	hypothetical prote		
28	165.5	14.0	1015	2	JC5552	DNA topoisomerase		
29	165	14.0	1214	2	T10737	extensin-like cell		

C:Accession: G93065; A03747; S32993
R:Bankier, A.T.; Deininger, P.L.; Farrell, P.J.; Barrell, B.G.
Mol. Biol. Med. 1, 21-45, 1983
A:Title: Sequence analysis of the 17,166 bp EcoRI fragment C of B95-8 Epstein-Barr virus
A:Reference number: A93065; MUID:85035713; PMID:6092825
A:Accession: G93065
A:Molecule type: DNA
A:Residues: 1-3149
A:Cross-references: EMBL:V01555; NID:g59074; PIDN:CAA24839.1; PID:g1334853
R:Baer, R.; Bankier, A.T.; Biggin, M.D.; Deininger, P.L.; Farrell, P.J.; Gibson, T.J.; H
Nature 310, 207-211, 1984
A:Title: DNA sequence and expression of the B95-8 Epstein-Barr virus genome.
A:Reference number: A03794; MUID:84270667; PMID:6087149
A:Contents: annotation; protein coding region
C:Superfamily: human herpesvirus 4 BPLF1 protein

Query Match	14.9%	Score 175;	DB 1;	Length 3149;
Best Local Similarity	34.08;	Pred. No. 0.0039;		
Matches 48;	Conservative 19;	Mismatches 40;	Indels 34;	Gaps 6;
Qy 105 PPRGFFP----	VP-----PSRFSSAAAPAA-----	PPIAAEPAAPAAALTAPEVA 145		
Db 290 PPVMLPERRIIPNLRPLPSRFTSDFSAARYSPAKTNSPP--	SSPASAAPASAPAS 347			
Qy 146 AEPAGAPVAEPAAEAPVGAEPAAEAPVAAEAPVGEPAEAPSPAEPAATAKP-A 204				
Db 348 AAPASAAPASAPASAPASAPASAPASPPPLFIPIGLHTPGVPAPSTPPRSSGA 407				
Qy 205 APE-----	PHPSPS 213			
Db 408 APQTPKRKKGLGKDSPHKKPT 428				

RESULT 11
T38236
Hypothetical protein SPAC23A1.17 - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C:Accession: T38236
R:Murphy, L.; Harris, D.; Wood, V.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, February 1998
A:Reference number: 221780
A:Accession: T38236
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-1611 <MUR>
A:Cross-references: EMBL:AL021813; PIDN:CAAL16951.1; GSPDB:GNO0066; SPDB:SPAC23A1.17
A:Experimental source: strain 972h-; cosmid c23A1
C:Genetics:
A:Gene: SPDB:SPAC23A1.17
A:Map position: 1

RESULT 12
T01456
extensin homolog F2401.18 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 23-Mar-2001
C:Accession: T01456
R:Shinn, P.; Buehler, E.; Dewar, K.; Feng, J.; Kim, C.; Li, Y.; Sun, H.; Conway, A.;
eologis, A.; Ecker, J.R.
submitted to the EMBL Data Library, January 1998
A:Description: Genomic sequence for Arabidopsis thaliana BAC F2401.
A:Reference number: Z14211
A:Accession: T01456
A>Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-786 <SH1>
A:Cross-references: EMBL:AC003113; NID:g2689438; PID:g2781362
A:Experimental source: cultivar Columbia
C:Genetics:
A:Map position: 1
A:Note: F2401.18

Query Match	14.7%	Score 173.5	DB 2	Length 786	
Best Local Similarity	29.7%	Pred. No. 0.0013			
Matches	62	Conservative 15	Mismatches 97	Indels 35	Gaps
Qy	33	PLHPSLNPYGIKRLNPPLYYRVPNTVPSVGGNTYDTGLPS--YFWILTSPGFPVYVHI	90		
Db	511	PPSPSPPPY-IYSSPPVNVCP-PTTQSPPPPKYEQTSPREYYP-----SPSPYYQYT	564		
Qy	91	RGFP-----LATQLNVPLPRGF-----PVVPSRFRFSAAAAPAAPPIAAEPAAAA--	137		
Db	565	SSPPPTYYATQSPPPPPPTYYXAVQSPPPPPVYPPVVTASTPSPPPPVYTYIQSPPPP	624		
Qy	138	PLTATVYAAGPAAGACVAAEPAAEAPVGA-----EPAAEAPVAAEPAAEAPV	184		
Db	625	PVYYSPTVQSPPPPPPVYYPVVTQSPPPSPVYPPVVTQSPPPPPVYLPVTQSPPPSPV	684		
Qy	185	GVEPAAEPSPAEATAKPAEPHFSPS	213		
Db	685	YYPVAKSPPPPPSVYPPVVTQSP-PPPS	712		

```

RESULT 13
S52418
GTP-binding regulatory protein Gs alpha-XL chain - rat
N:Alternate names: G protein XL-alpha-s
C:Species: Rattus norvegicus (Norway rat)
C:Date: 14-Jul-1995 #sequence_revision 10-Nov-1995 #text_change 02-Feb-2001
C:Accession: S52418
R:Kehlenbach, R.H.; Matthey, J.; Huttner, W.B.
Nature 372, 804-809, 1994
A:Title: XL-alpha-s is a new type of G protein.
A:Reference number: S52418; MUID:95089824; PMID:7997272
A:Accession: S52418
A:Molecule type: mRNA
A:Residues: 1-846 <KEY>
A:Cross-references: EMBL:X80407; NID:g642267; PID:CA558866.1; PID:g642268
R:Kehlenbach, R.H.; Matthey, J.; Huttner, W.B.
Nature 375, 253, 1995
A:Title: Correction: Xlalphas is a new type of G protein.

```

Query Match 14.6%; Score 172.5; DB 2; Length 846;
Best Local Similarity 33.0%; Pred. No. 0.0016;
Matches 77; Conservative 15; Mismatches 88; Indels 53; Gaps 13;
Qy 24 GEDD----NDGDHPLHPSINIPYGTIRNL-----PPLYRPNVTNTPS-YPGNYYTDT 70

Db 16 GVVDDTAVNMDSPPI-ASGPPPIEVSGAPVKSEHAKRPLERQAAETGNSPTSSTTAEPA 74
QY 71 GLPSYPWILTSFGFVYVHIGFPLATQ-----LNVPPLPRGFPFVPP 114
Db 75 KVPISLERGEESTQETVTHIKPAPVAEGSTOSSKADPDOSATHAVLQIGPEVGGVTMTPT 134
QY 115 SRFFSAAAPAAPIAAEP-AAAAPLTATPVAA-----EPAAGAPVAAEPAEAPVGAEP 169
Db 135 D--LPASDAGPDVRAEPDGGTAP--ATPAESEDNRPEAAAA--AAEPAEAP--AAEPA 186
QY 170 AE--APVAAEPAEAE-----PVGVEPAEPEPSAPAEPAATAKPAAPPEHPSP 212
Db 187 AEPAAEPAEPAEAEAPVDPTEAESAGSAGVDPDTEEPAAAAAATPAEPAARAAP 239
RESULT 14
G87675
arylesterase-related protein [imported] - Caulobacter crescentus
C:Species: Caulobacter crescentus
C>Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C:Accession: G87675
R.Nierman, W.C.: Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
n, J.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A:Title: Complete Genome Sequence of Caulobacter crescentus.
A:Reference number: A87249; MUID:21173698; PMID:11259647
A:Accession: G87675
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-438 <STO>
A:Cross-references: GB:AE005673; NID:g13425157; PIDN:AAK25403.1; GSPDB:GN00148
C:Genetics:
A:Gene: CC3441

Query Match 14.6%; Score 171.5; DB 2; Length 438;
Best Local Similarity 33.3%; Pred. No. 0.00095;
Matches 68; Conservative 14; Mismatches 59; Indels 63; Gaps 11;
QY 64 GNTYTDGLPSYP-----WILTSFGF-----PYVHIGFPLATQNLNVPPLPRGFPFVPP 114
Db 219 GRKYSADIKIYPNNAHLIDEPNTMKILDVIAWLGKGTQAPAAKVEPAPA----- 270
QY 115 SRFFSAAAPAA-APPIAAEPAEAAAAPLTATPV-AAEPAA-----GAPVAAEPA-----A 160
Db 271 ----PAAPAPAPAKAPEPAANAAP---EPVKAEPASPDKAKAPKADAKPKATA 323
QY 161 EAPVG-----ABEPAEAPV--AAEPAEAPVGVPEPAEPEPSAE----- 197
Db 324 KAPVAKAAPKAKAPAAEAPAKSAAPKAPAPAKAAAAAKPKAEAKPKTAAKAPAAE 383
QY 198 --PATAKPAAPPEHPSPSLEQANQ 219
Db 384 TAPAAKPAAPKAAKAAKPAKTATK 407

RESULT 15
T17547
proline-rich protein A57R - Chlorella virus PBCV-1
C:Species: Chlorella virus PBCV-1
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T17547
R.Graves, M.V.; Van Etten, J.L.
submitted to the EMBL Data Library, May 1999
A:Reference number: Z18806
A:Accession: T17547
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-544 <GRA>
A:Cross-references: EMBL:U42580; NID:g4028896; PIDN:AAC96425.1
A:Experimental source: specific host Chlorella strain NC64A
C:Genetics:

A:Gene: A57R
Query Match 14.5%; Score 171; DB 2; Length 544;
Best Local Similarity 27.8%; Pred. No. 0.0013;
Matches 58; Conservative 19; Mismatches 106; Indels 26; Gaps 4;
QY 2 KLLIWACIVCAFAKRRRF---PFIGEDDDNDGHLPLHPSLNIPYGINLPPPLYRPPVNT 58
Db 4 RVIFFLTLVTTAFARPEFTTNQIVTTQTVVNSPKPPSPKPPSPKPPSPKPPSP 63
QY 59 VPSYPGNTYTDGLPSYPWILTSFGFVYVHIGFPLATQNLNVPPLPRGFPFVPPSRFF 118
Db 64 KPSPKPPSPKPPSPKPP-----SP-----KPPSPKPPSP-KPPSPK 100
QY 119 SAAAPAAAPPIAAEPAEAAAAPLTATPVAAEPAAGAPVAAEPAEPAEPAEPAEPAE 178
Db 101 PSPKPPSPKPPSPKPPSPKPPSPKPPSPKPPSPKPPSPKPPSPKPPSPKPP 178
QY 179 AAEPVGVPEPAEPEPSAPAEPAATAKPAAPE 207
Db 161 PAPEPPALEPPAPEPPAPEPPAPEPPAPEPPAPEPPAPEPPAPEPPAPEPP 160

Search completed: February 12, 2003, 13:48:40
Job time : 51 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2003, 11:53:54 ; Search time 38 Seconds

(without alignments)
239.035 Million cell updates/sec

Title: US-09-923-236-2

Perfect score: 1178
Sequence: 1 MKLLWACIVCFARRRF.....TAKPAAPHPSPSLEQANQ 219

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	180	15.3	555	1	GPI_CHLRE
2	175	14.9	1054	1	IF2_STIAU
3	175	14.9	3149	1	TEGU_EBV
4	174	14.8	518	1	TPM4_DROME
5	170	14.4	3176	1	CA36_HUMAN
6	169	14.3	299	1	R122_DROME
7	169	14.3	865	1	CPN_DROME
8	167.5	14.2	450	1	CYL_PARDE
9	166	14.1	810	1	NFM_BOVIN
10	165.5	14.0	439	1	XP2_XENLA
11	160.5	13.6	992	1	EBN6_EBV
12	158	13.4	1274	1	ENAM_MOUSE
13	154	13.1	353	1	CCPA_ACERY
14	154	13.1	4499	1	DHA_CHLRE
15	152	12.9	721	1	YK82_MYCTU
16	151	12.8	88	1	H82_NEIGO
17	151	12.8	211	1	TUB8_SOLTU
18	150	12.7	352	1	ALGP_PSEAE
19	148.5	12.6	464	1	S3A2_HUMAN
20	147	12.5	1386	1	ZAP3_MOUSE
21	146.5	12.4	477	1	MA2_MOUSE
22	145.5	12.4	1565	1	PAC_STRMU
23	145	12.3	475	1	S3A2_MOUSE
24	145	12.3	2167	1	SHK1_RAT
25	144.5	12.3	477	1	MA2_HUMAN
26	142.5	12.1	497	1	WAS2_HUMAN
27	142.5	12.1	1083	1	T2D3_HUMAN
28	142	12.1	2161	1	SHK1_HUMAN
29	141.5	12.0	316	1	CDNC_HUMAN
30	141.5	12.0	465	1	FXD1_HUMAN
31	140.5	11.9	852	1	WS14_HUMAN
32	140	11.9	3421	1	TEGU_HSVB
33	139.5	11.8	397	1	SEPL_MOUSE

34 138.5 11.8 206 1 PRP1_MEDTR
35 138.5 11.8 1183 1 DRPL_RAT
36 138 11.7 376 1 FXL2_HUMAN
37 138 11.7 2517 1 NCR2_HUMAN
38 137 11.6 236 1 PRP_MEDSA
39 136 11.5 620 1 EXTN_TOBAC
40 136 11.5 1874 1 POLR_KYMUJ
41 135.5 11.5 817 1 VRP1_YEAST
42 135 11.5 346 1 PRF1_LYCES
43 134.5 11.4 1181 1 MYSC_DICDI
44 134 11.4 534 1 APG_ATH
45 133.5 11.3 1185 1 DRPL_HUMAN

O43564 medicago tr
P54258 rattus norv
P58012 homo sapien
Q9y618 h nuclear r
Q40358 medicago sa
P13983 nicotiana t
P36304 kennedyva ye
P37370 saccharomyc
Q00451 lycopersico
P42522 dictyostell
P40602 arabidopsis
P54259 homo sapien

ALIGNMENTS

RESULT 1

GPI_CHLRE STANDARD; PRT; 555 AA.

AC Q9FTQ6: Q03927; DT 15-JUN-2002 (rel. 41, Created)

DT 15-JUN-2002 (rel. 41, Last sequence update)

DT 15-JUN-2002 (rel. 41, Last annotation update)

DE Vegetative cell wall protein gpl precursor (Hydroxyproline-rich

glycoprotein 1).

GN Gpl

OS Chlamydomonas reinhardtii.

OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;

OC Chlamydomonadaceae; Chlamydomonas.

ON NCBI_TaxID=3055;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=21159092; PubMed=11258910;

RA Ferris P.J.; Woessner J.P.; Waffenschmidt S.; Kilz S.; Drees J.,

RA Goodenough U.W.;

RT "Glycosylated polyproline II rods-with-kinks as a structural motif in

plant hydroxyproline-rich glycoproteins.";

RL Biochemistry 40:2978-2987(2001).

RN [2]

RP PARTIAL PRELIMINARY SEQUENCE FROM N.A.

RX MEDLINE=91017504; PubMed=1699225;

RA Adair W.S.; Apt K.E.;

RT "Cell wall regeneration in Chlamydomonas: accumulation of mRNAs

encoding cell wall hydroxyproline-rich glycoproteins.";

Proc. Natl. Acad. Sci. U.S.A. 87:7355-7359(1990).

-!- FUNCTION: Major component of the outer cell wall w6 (crystalline)

layer

-!- SUBUNIT: Associates with GP2 and GP3.

-!- PTM: N-glycosylated and O-glycosylated.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; AF309494; AAG45420.1; -

DR EMBL; M58496; AAG69706.1; ALT_SEQ.

DR GlycoSuiteDB; Q9FPO6; -

DR InterPro; IPR002965; P_rich_extensin.

DR InterPro; IPR003882; PstII_extensin.

DR PRINTS; PR01217; PRICHTEXTENS.

DR PRINTS; PR01218; PSTLEXTENSIN.

KW Glycoprotein; Repeat; Signal.

FT SIGNAL 1 29 POTENTIAL.

FT CHAIN 30 555 VEGETATIVE CELL WALL PROTEIN GPI.

FT DOMAIN 40 339 49 X 5 AA APPROXIMATE PPSPX REPEATS.

FT DOMAIN 259 279 POLY-PRO.

FT CARBOHYD 399 399 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 455 455 N-LINKED (GLCNAC. . .) (POTENTIAL).

RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=Fibroblast;
RX MEDLINE=90151612; PubMed=1689238;
RA Chu M.-L., Zhang R.-Z., Pan T.-C., Stokes D., Conway D., Kuo H.-J.,
RT "Mosaic structure of globular domains in the human type VI collagen
RA alpha 3 chain: similarity to von Willebrand factor, fibronectin,
RT actin, salivary proteins and aprotinin type protease inhibitors.";
RL EMBO J. 9:385-393(1990).
RN [2]
RP REVISIONS.
RA Chu M.-L.;
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 2038-2373 FROM N.A.
RX MEDLINE=89066644; PubMed=3198591;
RA Chu M.-L., Conway D., Pan T.-C., Baldwin C., Mann K., Deutzmann R.,
RA Timpl R.;
RT "Amino acid sequence of the triple-helical domain of human collagen
RT type VI.";
RL J. Biol. Chem. 263:18601-18606(1988).
RN [4]
RP SEQUENCE OF 2092-2157 FROM N.A.
RX MEDLINE=88029444; PubMed=3665927;
RA Chu M.-L., Mann K., Deutzmann R., Pribula-Conway D.,
RA Hsu-Chen C.-C., Bernard M.P., Timpl R.;
RT "Characterization of three constituent chains of collagen type VI by
RT peptide sequences and cDNA clones.";
RL Eur. J. Biochem. 168:309-317(1987).
RN [5]
RP SEQUENCE OF 2092-2151 FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=88161046; PubMed=3348212;
RA Weil D., Mattei M.-G., Passage E., van Cong N., Pribula-Conway D.,
RA Mann K., Deutzmann R., Timpl R., Chu M.-L.;
RT "Cloning and chromosomal localization of human genes encoding the
RT three chains of type VI collagen.";
RL Am. J. Hum. Genet. 42:435-445(1988).
RN [6]
RP SEQUENCE OF 32-236 FROM N.A., AND ALTERNATIVE SPLICING.
RX MEDLINE=93054780; PubMed=1339440;
RA Zanussi S., Doliara K., Segat D., Bonaldo P., Colombatti A.;
RT "The human type VI collagen gene. mRNA and protein variants of the
RT alpha 3 chain generated by alternative splicing of an additional 5-end
RT exon.";
RL J. Biol. Chem. 267:24082-24089(1992).
RN [7]
RP X-RAY CRYSTALLOGRAPHY (1.6 ANGSTROMS) OF 3107-3164.
RX MEDLINE=95182468; PubMed=7533217;
RA Arnoux B., Merigau K., Saludjian P., Norris F., Norris K., Bjoern S.,
RA Olsen O., Petersen L., Ducruix A.;
RT "The 1.6 A structure of Kunitz-type domain from the alpha 3 chain of
RT human type VI collagen.";
RL J. Mol. Biol. 246:609-617(1995).
RN [8]
RP STRUCTURE BY NMR OF 3102-3164.
RX MEDLINE=96398604; PubMed=8805527;
RA Zwickstetter M., Czisch M., Mayer U., Chu M.-L., Zinth W., Timpl R.,
RA Holak T.A.;
RT "Structure and multiple conformations of the kunitz-type domain from
RT human type VI collagen alpha3(VI) chain in solution.";
RL Structure 4:195-209(1996).
RN [9]
RP STRUCTURE BY NMR OF 3107-3164.
RX MEDLINE=97410331; PubMed=9265624;
RA Soerensen M.D., Bjoern S., Norris K., Olsen O., Petersen L.,
RA James T.L., Led J.J.;
RT "Solution structure and backbone dynamics of the human alpha3-chain
RT type VI collagen C-terminal Kunitz domain.";
RL Biochemistry 36:10439-10450(1997).
RN [10]
RP VARIANT BM GLU-1679, AND VARIANT HIS-2831.
RX MEDLINE=98204804; PubMed=9536084;

RA Pan T.-C., Zhang R.-Z., Pericak-Vance M.A., Tandan R., Fries T.,
RA Stajich J.M., Viles K., Vance J.M., Chu M.-L., Speer M.C.;
RT "Missense mutation in a von Willebrand factor type A domain of the
RT alpha 3(VI) collagen gene (COL6A3) in a family with Bethlem
RT myopathy.";
RL Hum. Mol. Genet. 7:807-812(1998).
CC -!- FUNCTION: COLLAGEN VI ACTS AS A CELL-BINDING PROTEIN.
CC -!- SUBUNIT: TRIMERS COMPOSED OF THREE DIFFERENT CHAINS: ALPHA 1(VI),
CC ALPHA 2(VI), AND ALPHA 3(VI).
CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; 1 (SHOWN HERE) AND 2; ARE
CC PRODUCED BY ALTERNATIVE SPLICING.
CC -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC -!- DISEASE: DEFECTS IN COL6A3 ARE A CAUSE OF BETHLEM MYOPATHY (BM).
CC BM IS A RARE AUTOSOMAL DOMINANT PROXIMAL MYOPATHY CHARACTERIZED BY
CC EARLY CHILDHOOD ONSET (COMPLETE PENETRANCE BY THE AGE OF 5) AND
CC JOINT CONTRACTURES MOST FREQUENTLY AFFECTING THE ELBOWS AND
CC ANKLES.
CC -!- SIMILARITY: CONTAINS 1 BPTI/KUNITZ INHIBITOR DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 FIBRONECTIN TYPE III-LIKE DOMAIN.
CC -!- SIMILARITY: CONTAINS 12 VWFA DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X52022; CAA36267.1; -;
DR EMBL; X06196; CAA29557.1; -;
DR EMBL; M20778; -: NOT_ANNOTATED_CDS.
DR EMBL; M27449; AAA52057.1; -;
DR EMBL; S49432; AAB24261.1; -;
DR PIR; C31952; C31952.
DR PDB; 1KNT; 01-NOV-94.
DR PDB; 2KNT; 15-MAY-97.
DR PDB; 1KUN; 12-NOV-97.
DR Genew; HGNC:2213; COL6A3.
DR MIM; 120250; -;
DR MIM; 158810; -;
DR InterPro; IPR000087; Collagen.
DR InterPro; IPR003961; FN_III.
DR InterPro; IPR002223; Kunitz_BPTI.
DR InterPro; IPR002035; VWF_A.
DR Pfam; PF00014; Kunitz_BPTI; 1.
DR Pfam; PF00092; vwa; 11.
DR Pfam; PF01391; Collagen; 5.
DR PRINTS; PR00759; BASICPTASE.
DR PRINTS; PR00453; VWFADOMAIN.
DR ProDom; PD000222; Kunitz_BPTI; 1.
DR SMART; SM00060; FN3; 1.
DR SMART; SM00131; KU; 1.
DR SMART; SM00327; VWF; 12.
DR PROSITE; PS00280; BPTI_KUNITZ_1; 1.
DR PROSITE; PS0279; BPTI_KUNITZ_2; 1.
DR PROSITE; PS0234; VWFA; 12.
KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW Glycoprotein; Cell adhesion; Serine protease inhibitor; Collagen;
KW Signal; 3D-structure; Disease mutation; Polymorphism;
KW Alternative splicing.
FT SIGNAL 1 25
FT CHAIN 26 3176
FT COLLAGEN ALPHA 3(VI) CHAIN.
FT DOMAIN 26 2038
FT NONHELICAL REGION.
FT DOMAIN 2039 2375
FT TRIPLE-HELICAL REGION.
FT DOMAIN 2376 3176
FT NONHELICAL REGION.
FT DOMAIN 39 213
FT VWFA 1.
FT DOMAIN 242 419
FT VWFA 2.
FT DOMAIN 445 620
FT VWFA 3.
FT DOMAIN 639 816
FT VWFA 4.
FT DOMAIN 837 1009
FT VWFA 5.
FT DOMAIN 1029 1205
FT VWFA 6.


```

FT DOMAIN 1233 1404 VWFA 7.
FT DOMAIN 1436 1609 VWFA 8.
FT DOMAIN 1639 1812 VWFA 9.
FT DOMAIN 1838 2024 VWFA 10.
FT DOMAIN 2402 2581 VWFA 11.
FT DOMAIN 2619 2815 VWFA 12.
FT DOMAIN 2987 3076 FIBRONECTIN TYPE-III.
FT DOMAIN 3107 3176 BPTI/KUNITZ INHIBITOR.
FT SITE 2040 2042 CELL ATTACHMENT SITE.
FT SITE 2136 2138 CELL ATTACHMENT SITE.
FT SITE 2148 2150 CELL ATTACHMENT SITE.
FT SITE 2154 2156 CELL ATTACHMENT SITE.
FT SITE 2370 2372 CELL ATTACHMENT SITE.
FT ACT_SITE 3121 3122 REACTIVE BOND.
FT DISULFID 3111 3161
FT DISULFID 3120 3144
FT DISULFID 3136 3157
FT CARBOHYD 202 202
FT CARBOHYD 251 251
FT CARBOHYD 2079 2079
FT CARBOHYD 2331 2331
FT CARBOHYD 2558 2558
FT CARBOHYD 2677 2677
FT CARBOHYD 2861 2861
FT CARBOHYD 3036 3036
FT VARSPLIC 32 236
FT VARIANT 1679 1679
FT VARIANT 2831 2831
FT CONFLICT 127 128
FT CONFLICT 137 137
FT CONFLICT 2157 2157
SQ SEQUENCE 3176 AA; 343548 MW; 5DF82563229BB2DD CRC64;

Query Match 14.4%; Score 170; DB 1; Length 3176;
Best Local Similarity 31.8%; Pred. No. 0.013;
Matches 57; Conservative 16; Mismatches 56; Indels 50; Gaps 9;

QY 35 HPSLNIPYGIKRLP--DPLXYRPVNTVPSYPGNYITDTGLPSYPIWLTSPGFVPVYHIR 91
Db 2853 HQQVNPVNTSSPSNPVTITKPTVTT-----KPVTTITKP-----VTITTKP----- 2896
QY 92 GFPLATQLNVLPLPRGFFVFPSPRFFSAAAPAA-APPIAAEPAAAPALPTATPVAEPA 150
Db 2897 ----VTIINQPSVKP-----AAAKPAPAKPVAAKPVATKATVRP----- 2932
QY 151 GAPVAEPAEPAEA-PVGAEPAAEAPVAEPAEAPVGVPEPAEAEPSPAEPATAKPAPEP 208
Db 2933 --FVAVKPKATAKPVAAKPAAPVRRPPA---AAAKPVATKPEVPRPQAAKPAATKPAATKP 2986

RESULT 6
RL22_DROME
ID RL22_DROME STANDARD; PRT; 299 AA.
AC P50887; Q9V3X9;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE 60S ribosomal protein L22.
GN RPL22 OR EG:BACRI9J1.4 OR CG7434.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_taxid=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA Glover C.V.C., Bidwai A.P., Zhao W.F.;
RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Oregon-R;

```

```

RX MEDLINE=20196011; PubMed=10731137;
RA Benos P.V., Gatt M.K., Ashburner M., Murphy L., Harris D.,
RA Barrell B.G., Ferraz C., Vidal S., Brun C., Demailles J., Cadieu E.,
RA Dreano S., Gloux S., Lelaure V., Mottier S., Galibert F., Borkova D.,
RA Papagiannakis G., Spanos L., Cox S., Madueno E., Bolshakov S.,
RA Modolell J., Peter A., Schoettler P., Werner M., Mourikioti F.,
RA Beinert N., Dowe G., Schaefer U., Jaekle H., Bucheton A.,
RA Callister D.M., Campbell L.A., Darlamitsou A., Henderson N.S.,
RA McMillan P.J., Salles C., Tait E.A., Valenti P., Saunders R.D.C.,
RA Glover D.M.;
RT "From sequence to chromosome: the tip of the X chromosome of D.
RT melanogaster.";
RL Science 287:2220-2222(2000).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkely;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Ananidis P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abrial J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Bottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fosler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Hostin D.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
RN [1]
RP SIMILARITY: BELONGS TO THE L22E FAMILY OF RIBOSOMAL PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; U42587; AAB17433.1; -
CC EMBL; AL132792; CAB60023.1; -
CC EMBL; AE003418; AAF45546.1; -
CC FlyBase; FBgn0015288; RPL22.
CC InterPro; IPR002671; Ribosomal_L22e.
CC Pfam; PF01776; Ribosomal_L22e; 1.
CC ProDom; PD007306; Ribosomal_L22e; 1.

```

```
KW Ribosomal protein..
FT DOMAIN 24 31 POLY-ALA.
FT DOMAIN 46 50 POLY-ALA.
FT DOMAIN 65 70 POLY-ALA.
FT DOMAIN 93 98 POLY-ALA.
FT DOMAIN 103 112 POLY-ALA.
FT DOMAIN 136 152 POLY-ALA.
FT DOMAIN 185 188 POLY-LYS.
FT DOMAIN 292 299 ASP/GLU-RICH (HIGHLY ACIDIC).
SQ SEQUENCE 299 AA; 30610 MW; 46A99005610E4EB0 CRC64;

Query Match 14.3%; Score 169; DB 1; Length 299;
Best Local Similarity 44.1%; Pred. No. 0.0021;
Matches 49; Conservative 6; Mismatches 34; Indels 22; Gaps 5;

QY 121 AAAPAAPPTAAEPAAAPLITAT-----PVAAEPAAAGAPVAAEPAAEPVCAEPAAEAP 173
DB 65 AAAAAKPAAPAAKPAAPKADGAKGAPAAAAPKDKAKAAPAAAPAKAPAKAAATP 124
QY 174 VAAEP-----AEAPGVGEPAEPEPAETAKPAPEPH-----PSPS 213
DB 125 AAAPPAKAAKAAKAAAPAAAPAA-----PAAAPAVAKP-APKPKAAAPAPS 171

RESULT 7
CPN_DROME STANDARD; PRT; 865 AA.
AC Q02910;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DE 01-FEB-1994 (Rel. 28, Last annotation update)
DE Calphotin.
DE CPN OR CAP.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Canton-S;
RX MEDLINE=93165729; PubMed=8094559;
RA Martin J.H., Benzer S., Rudnicka M., Miller C.A.;
RT "Calphotin: a Drosophila photoreceptor cell calcium-binding protein.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:1531-1535(1993).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN-Canton-S;
RX MEDLINE=93165730; PubMed=8434015;
RA Ballinger D.G., Xue N., Harshman K.D.;
RT "A Drosophila photoreceptor cell-specific protein, calphotin, binds
calcium and contains a leucine zipper.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:1536-1540(1993).
CC -!- FUNCTION: MIGHT FUNCTION AS A CALCIUM-SEQUESTERING "SPONGE" TO
REGULATE THE AMOUNT OF FREE CYTOPLASMIC CALCIUM. IT BINDS 0.3 MOL
OF CA+2 PER MOL OF PROTEIN.
CC -!- SUBUNIT: HOMODIMER (PROBABLE).
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC; HYPODENSE COMPARTMENT.
CC -!- TISSUE SPECIFICITY: SOMA AND AXONS OF PHOTORECEPTOR CELLS OF
COMPOUND EYES AND OCELLI.
CC -!- DEVELOPMENTAL STAGE: EXPRESSED EARLY IN PHOTORECEPTOR CELL
DEVELOPMENT.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; L02111; AAA28405.1; -
EMBL; L05080; AAA28420.1; -
```

```
DR PIR; A47282; A47282.
DR FlyBase; FBgn0010218; Cpn.
KW Calcium-binding.
FT CONFLICT 36 A -> AVAPAVVA (IN REF. 2).
FT CONFLICT 43 I -> T (IN REF. 2).
FT CONFLICT 64 I -> V (IN REF. 2).
FT CONFLICT 76 T -> A (IN REF. 2).
FT CONFLICT 100 P -> PP (IN REF. 2).
FT CONFLICT 126 VQ -> AP (IN REF. 2).
FT CONFLICT 154 I -> V (IN REF. 2).
FT CONFLICT 160 S -> T (IN REF. 2).
FT CONFLICT 534 A -> E (IN REF. 2).
FT CONFLICT 699 I -> T (IN REF. 2).
FT CONFLICT 703 V -> L (IN REF. 2).
FT CONFLICT 721 D -> E (IN REF. 2).
SQ SEQUENCE 865 AA; 84781 MW; 2110417E080E7CFC CRC64;

Query Match 14.3%; Score 169; DB 1; Length 865;
Best Local Similarity 29.5%; Pred. No. 0.005;
Matches 70; Conservative 14; Mismatches 75; Indels 78; Gaps 11;

QY 55 PVNTVPSPGNTYTDGLPSYDWLLTSPGFPYVYHTRGFPLAT---QLNVPLPLPRGFPP 111
DB 42 PIAVTPVAPPTLASV---OPATVTIPAPATAAASVTGASVAPVVAAPTTPAASPV 97
QY 112 VPP---SRFFSAAAAAPAPPIAAEPA-----AAAPLTAT-PVAAE---PAA----- 150
DB 98 STPVAVAQIPVAVSAPVAPVPAATPTPVVQIPVAAAPVIATTPVVAASAPTPAAVTPVISPV 157
QY 151 -----GAPVAAEPA-----AEAPVGAE-----PAAEAPVA 175
DB 158 IASPPVVPANTTVPVAAAPVAAVPAVPAVPAVPAVPAVPAVPAVPAVPAVPAVPAVPAV 217
QY 176 AEPAEAPV-----GVEPAEEPSAEATAPAAPEPHPS---PSLEOA 217
DB 218 TIPECVAPLPEVSVVATKPLAAAEPPVVVAPPATETPVVAPAAASHVSVAPAVETA 274

RESULT 8
CYL_PARDE STANDARD; PRT; 450 AA.
AC P13627;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Cytochrome c1 precursor.
GN PETC.
OS Paracoccus denitrificans.
OC Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;
OC Paracoccus.
OX NCBI_TaxID=266;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88007612; PubMed=2820981;
RA Kuroski B., Ludwig B.;
RT "The genes of the Paracoccus denitrificans bcl complex. Nucleotide
sequence and homologies between bacterial and mitochondrial
subunits.";
RL J. Biol. Chem. 262:13805-13811(1987).
CC -!- FUNCTION: COMPONENT OF THE UBIQUINOL-CYTOCHROME C REDUCTASE
COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH IS A
RESPIRATORY CHAIN THAT GENERATES AN ELECTROCHEMICAL POTENTIAL
COUPLED TO ATP SYNTHESIS. C1 FUNCTIONS AS AN ELECTRON DONOR TO
CYTOCHROME C.
CC -!- SUBUNIT: THE MAIN SUBUNITS OF COMPLEX B-C1 ARE: CYTOCHROME B,
CYTOCHROME C1 AND THE RIESKE PROTEIN.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; L02111; AAA28405.1; -
EMBL; L05080; AAA28420.1; -
```

```

CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M17522; AAA25573.1; -.
DR EMBL: X05799; CAA29245.1; -.
DR PIR: C29413; C29413.
DR InterPro: IPR000345; CytC_heme_bind.
DR InterPro: IPR002326; Cyt_C1.
DR Pfam: PF02167; Cytochrome.C1; 1.
DR PRINTS: PR00603; CYTOCHROME.C1.
DR PROSITE: PS00190; CYTOCHROME_C; 1.
KW Electron transport; Respiratory chain; Oxidative phosphorylation;
KW Heme; Transmembrane; Signal.
FT SIGNAL 1 21
FT CHAIN 22 450 CYTOCHROME C1.
FT BINDING 245 245 HEME (COVALENT).
FT BINDING 248 248 HEME (COVALENT).
FT METAL 249 249 IRON (HEME AXIAL LIGAND).
FT METAL 373 373 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
FT TRANSMEM 421 435 ANCHORS TO THE MEMBRANE (POTENTIAL).
SQ SEQUENCE 450 AA; 46874 MW; C1EA8B6B48FE7603 CRC64;

Query Match 14.28; Score 167.5; DB 1; Length 450;
Best Local Similarity 42.68; Pred. No. 0.0035;
Matches 58; Conservative 8; Mismatches 35; Indels 35; Gaps 8;

QY 114 PSRFSSAAAAPPAAPPIAAEPAAA--APLTATPVAA-----EPAAGAPVA-----AEPAA 159
D 48 PAAADTAPAAEAADPEAAEEAAGEAEVTEPATETPAEPADEFAAETPEPDAEPA 107
QY 160 -----AEAPVGAEPAAEPVAAEPAAEPVGEPAEPAEPSPAE--ATAKPA 205
D 108 ABEAQATTEAPAEPAEPAEPAEPAEPAEPAEPAEPAEPAEPAEPAEPAEPAE 164
QY 206 PPP-----HPSPSLEQA 217
D 165 EEPATEEPAEPAEAA 180

RESULT 9
ID NFM_BOVIN STANDARD; PRT; 810 AA.
AC O77788;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Neurofilament triplet M protein (160 kDa neurofilament protein)
DE (Neurofilament medium polypeptide) (NF-M) (fragment).
GN NEF3 OR NFM OR NFM.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RA Hill W.D., Zhang L., Balin B.J., Sprinkle T.J., Spicer K.,
RA Gearhart D.A.;
RT "The bovine neurofilament M subunit has a novel set of KSP repeats
RT normally restricted to NF-H.;"
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,
CC AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.
CC -!- PTM: THERE ARE A NUMBER OF REPEATS OF THE TRIPEPTIDE K-S-P. NFM IS
CC PHOSPHORYLATED ON A NUMBER OF THE SERINES IN THIS MOTIF. IT IS
CC THOUGHT THAT PHOSPHORYLATION OF NFM RESULTS IN THE FORMATION OF
CC OF AXONAL CALIBER (BY SIMILARITY).
CC -!- PTM: PHOSPHORYLATION SEEMS TO PLAY A MAJOR ROLE IN THE FUNCTIONING
CC OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H). THE
CC LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND
CC COINCIDENT WITH A CHANGE IN THE NEUROFILAMENT FUNCTION (BY
CC SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.

-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF091342; AAC36357.1; -.
DR InterPro: IPR001664; IF.
DR Pfam: PF00038; filament; 1.
DR PROSITE: PS00226; IF; 1.
KW Intermediate filament; Coiled coil; Neurone; Phosphorylation.
FT NON_TER 1 1
FT DOMAIN <1 296 ROD.
FT DOMAIN 297 810 TAIL.
FT DOMAIN <1 20 COIL 1A.
FT DOMAIN 31 33 LINKER 1.
FT DOMAIN 34 132 COIL 1B.
FT DOMAIN 133 149 LINKER 12.
FT DOMAIN 150 171 COIL 2A.
FT DOMAIN 172 175 LINKER 2.
FT DOMAIN 176 296 COIL 2B.
FT DOMAIN 503 582 8 X 13 AA TANDEM REPEATS.
SQ SEQUENCE 810 AA; 90799 MW; B8477D85560AC3F6 CRC64;

Query Match 14.18; Score 166; DB 1; Length 810;
Best Local Similarity 35.18; Pred. No. 0.0069;
Matches 34; Conservative 12; Mismatches 51; Indels 0; Gaps 0;

QY 112 VPPSRFSSAAAAPPAAPPIAAEPAAAAPLTATPVAAEPAAAGAPVAAEPAAEPVGAEPAAE 171
D 484 VAPKEELAAEAKVEKEKAKSPVAKSPTKSPAKSPAKSPAKSPAKSPAKSPAKSPAKSPAKSP 543
QY 172 APVAAEPAAEPVGVBPAAEPSPAEPAEPATAKPAAP 208
D 544 SPTAKSPAKSPAKSPAKSPAKSPAKSPAKSPAKSPAKSPAKSPAKSPAKSPAKSPAKSPAKSP 580

RESULT 10
ID XP2_XENLA STANDARD; PRT; 439 AA.
AC P17437; Q08944;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Skin secretory protein XP2 precursor (APEG protein).
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE OF 1-25 AND 344-439 FROM N.A. (ISOFORM 2).
RA Hausser F., Roeben C., Hoffmann W.;
RT "xp2, a new member of the p-domain peptide family of potential growth
RT factors, is synthesized in xenopus laevis skin.;"
RL J. Biol. Chem. 267:14451-14455(1992).
RN [2]
RP SEQUENCE OF 3-439 FROM N.A. (ISOFORM 1).
RC TISSUE-Skin;
RX MEDLINE=90127399; PubMed=2298293;
RA Gmachl M., Berger H., Thalhammer J., Kreil G.;
RT "Dermal glands of Xenopus laevis contain a polypeptide with a highly
RT repetitive amino acid sequence.;"
RL FEBS Lett. 260:145-148(1990).
CC -!- FUNCTION: MAY ACT AS A GROWTH FACTOR IN THE GERMINAL LAYER OF THE
CC EPIDERMIS. MAY ALSO BE INVOLVED IN GROWTH OF REGENERATING GLANDS
CC AND IN PROTECTION OF THE SKIN FROM THE EXTERNAL ENVIRONMENT.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- ALTERNATIVE PRODUCTS: 2 isoforms; 1/APEG (shown here) and 2/xp2;

```

```
CC are produced by alternative splicing.
CC -!- TISSUE SPECIFICITY: SKIN.
CC -!- SIMILARITY: CONTAINS 2 P-TYPE (TREFOL) DOMAINS.
CC -!- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN FROM POSITION 392
CC ONWARD AND IS SHORTER (418 AA) DUE TO A FRAMESHIFT.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M90095; AAA50001.1; -
DR EMBL; X51394; CAA35759.1; ALT_FRAME.
DR PIR; S07498; SKLAG.
DR PIR; A37331; A37331.
DR HSP; P04155; IP52.
DR InterPro; IPR000519; P_trefol.
DR Pfam; PF00088; trefol; 2.
DR PRINTS; PRO0680; TREFOL.
DR SMART; SM00018; P; 2.
DR PROSITE; PS00025; P_TREFOL; 2.
KW Signal; Growth factor; Alternative splicing; Amphibian skin; Repeat.
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 439 SKIN SECRETORY PROTEIN XP2.
FT MOD_RES 23 23 PYRROLIDONE CARBOXYLIC ACID (PROBABLE).
FT DOMAIN 26 343 33 X REPEATS OF G-[GE]-[AP](2,4)-A-E.
FT DOMAIN 350 391 P-TYPE 1.
FT DOMAIN 397 438 P-TYPE 2.
FT DISULFD 351 377 BY SIMILARITY.
FT DISULFD 371 388 BY SIMILARITY.
FT DISULFD 398 424 BY SIMILARITY.
FT DISULFD 408 423 BY SIMILARITY.
FT DISULFD 418 435 BY SIMILARITY.
FT VARSPLOT 26 343 MISSING (IN ISOFORM 2).
FT CONFLICT 3 3 H -> S (IN REF. 2).
FT CONFLICT 18 18 C -> W (IN REF. 2).
SQ SEQUENCE 439 AA; 41173 MW; 38C4A4B57CBAE778 CRC64;

Query Match 14.0%; Score 165.5; DB 1; Length 439;
Best Local Similarity 39.7%; Pred. No. 0.0044;
Matches 48; Conservative 6; Mismatches 56; Indels 11; Gaps 3;

QY 103 PLPPRGFFVPSRFFSAAAPAPPIAAEPAAAPLTATPVAAEPAG-APVAAEPAAE 161
Db 103 PAPAEGGAPAP-----AEGGAPAPAPAGEAPAPAPAGEAPAPAPAGEAPAGEAP 157

QY 162 APVGAEPAAEPVAAEPAAEPVGEPAEPPSPAE-----PATAPAPAPPHPSPSLEQ 216
Db 158 APAPAEVEAPAPAPAGEAPAPAPAGEAPAPAPAGEAPAPAPAGEAPAPAPAGE 217

QY 217 A 217
Db 218 A 218

RESULT 11
ENB6_EBV ID ENB6_EBV STANDARD; PRT; 992 AA.
AC P03204;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 01-FEB-1994 (Rel. 28, Last annotation update)
DE EBNA-6 nuclear protein (EBNA-3C) (EBNA-4B).
GN BEBF3-BERF4.
OS Epstein-Barr virus (strain B95-8) (Human herpesvirus 4).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OX Gammaherpesvirinae; Lymphocryptovirus.
OC NCBI_TaxId=10377;
RN [1]
```

```
RP SEQUENCE FROM N.A.
RX MEDLINE=84270667; PubMed=6087149;
RA Baer R., Bankier A.T., Biggin M.D., Deininger P.L., Farrell P.J.,
RA Gibson T.J., Hatfull G., Hudson G.S., Satchwell S.C., Sequin C.,
RA Tuffnell P.S., Barrell B.G.;
RT "DNA sequence and expression of the B95-8 Epstein-Barr virus genome.";
RL Nature 310:207-211(1984).
RN [2]
RP CHARACTERIZATION.
RX MEDLINE=88155772; PubMed=2831394;
RA Petti L., Sample J., Wang F., Kieff E.;
RT "A fifth Epstein-Barr virus nuclear protein (EBNA3C) is expressed in
RT latently infected growth-transformed lymphocytes.";
RL J. Virol. 62:1330-1338(1988).
RN [3]
RP SUBCELLULAR LOCATION.
RX MEDLINE=90266473; PubMed=2161150;
RA Petti L., Sample C., Kieff E.;
RT "Subnuclear localization and phosphorylation of Epstein-Barr virus
RT latent infection nuclear proteins.";
RL Virology 176:563-574(1990).
CC -!- FUNCTION: INVOLVED IN LATENT CYCLE.
CC -!- SUBCELLULAR LOCATION: NUCLEAR. ASSOCIATED WITH THE NUCLEAR MATRIX.
CC -!- SIMILARITY: SOME SIMILARITIES EXIST BETWEEN EBNA 4, 5, AND 6.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; V01555; CAA24859.1; -.
DR PIR; A03766; Q0BE25.
KW Nuclear protein; Repeat.
FT DOMAIN 74 80 POLY-ARG.
FT DOMAIN 551 610 10 X 5 AA TANDEM REPEATS.
FT DOMAIN 741 779 3 X 13 AA TANDEM REPEATS.
SQ SEQUENCE 992 AA; 109129 MW; 39BEAB9BC515BD84 CRC64;

Query Match 13.6%; Score 160.5; DB 1; Length 992;
Best Local Similarity 26.1%; Pred. No. 0.017;
Matches 62; Conservative 22; Mismatches 83; Indels 71; Gaps 8;

QY 33 PLHPS-LNIPYGIIRNLPPPLYRYEVNTVPSYGN-----TYTDT 70
Db 464 PAHPTVEMPNVILHPPPV-PKPVPVKPTPPPPRRRGACVYDDVIEIDVETEDS 522

QY 71 GLPSYFWILTSPGFVYVYHIRGFPLATQLNVPLPRGFPFVPPSRFFSAAAPAPPIA 130
Db 523 S-----SVSQPNKPKRKHODGFQSRGRQKRAAPTVSP-----SDTGPPVGP 568

QY 131 AEPAAPALATPVAAEPAPAGAPVAAEPAAEPVGAEPAAE-----APVAAEPAAEAP-- 183
Db 569 AGPPAAAGPPAAGPPAAGPPAAGPPAAGPPAAGPPAAGPPAAGPPAAGPPAAGPPA 628

QY 184 -----VGVEPAA-----EPSPAEPAEPATAPAPAPHPSPSL 214
Db 629 VRMFMRERQLPQSTGRKPOCFWEMRAGREITOMQEPSSHLSQATQPTTPRPSWAPSV 686

RESULT 12
ENAM_MOUSE ID ENAM_MOUSE STANDARD; PRT; 1274 AA.
AC O55196;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Enamelin precursor.
GN ENAM.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```


GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2003, 11:57:24 ; Search time 61 seconds
(without alignments)
739.743 Million cell updates/sec

Title: US-09-923-236-2

Perfect score: 1178
Sequence: 1 MKLLWACIVCVAFARRRF.....TAKPAPEHPSPSEQANO 219

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- SPTREMBL_21.*
- 1: sp.archaea.*
 - 2: sp.bacteria.*
 - 3: sp.fungi.*
 - 4: sp.human.*
 - 5: sp.invertebrate.*
 - 6: sp.mammal.*
 - 7: sp.mhc.*
 - 8: sp.organelle.*
 - 9: sp.phage.*
 - 10: sp.plant.*
 - 11: sp.rodent.*
 - 12: sp.virus.*
 - 13: sp.vertibrate.*
 - 14: sp.unclassified.*
 - 15: sp.virus.*
 - 16: sp.bacteriap.*
 - 17: sp.archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	359.5	30.5	344	11	Q9D5J3	Q9d5j3 mus musculus
2	224.5	19.1	3179	12	Q8V2A4	Q8v2a4 human herpe
3	221	18.8	1100	2	Q9X6J9	Q9x6j9 klebsiella
4	214.5	18.2	1480	10	Q9LIE8	Q9lie8 arabidopsis
5	214	18.2	3105	12	Q8UZ19	Q8uz19 cercopithec
6	206.5	17.5	607	5	Q71112	Q71112 babesia bov
7	199	16.9	391	5	Q26892	Q26892 trypanosoma
8	194.5	16.5	722	5	Q26893	Q26893 trypanosoma
9	193	16.4	3534	12	Q39266	Q39266 equine herp
10	188.5	16.0	1108	2	Q49542	Q49542 mycoplasma
11	188	16.0	839	16	Q9RX57	Q9rx57 deinococcus
12	186.5	15.8	395	16	Q9RV74	Q9rv74 deinococcus
13	186.5	15.8	616	11	Q99K31	Q99k31 mus musculus
14	183	15.5	744	10	Q65375	Q65375 arabidopsis
15	182	15.4	386	10	Q9FPQ5	Q9fpq5 chlamydomon
16	181	15.4	928	12	Q9IMX9	Q9imx9 cercopithec

17	181	15.4	2657	11	O88493	O88493 mus musculus
18	178.5	15.2	243	10	Q9XIV1	Q9xiv1 cucumis sat
19	177.5	15.1	598	16	O8VKN7	O8vkn7 mycobacteri
20	177	15.0	1315	10	Q9SPM0	Q9spm0 zea mays (m
21	176.5	15.0	3084	12	Q8UZ11	Q8uz11 pseudorabie
22	176	14.9	959	16	O8XRH0	O8xrh0 raistonia s
23	175	14.9	857	16	Q98IK2	Q98ik2 rhizobium l
24	174.5	14.8	1611	3	O42854	O42854 schizosacch
25	174	14.8	236	10	Q9LM00	Q9lm00 pinus taeda
26	174	14.8	503	12	Q39779	Q39779 equine herp
27	174	14.8	566	5	Q9VF97	Q9vf97 drosophila
28	173.5	14.7	786	10	O48809	O48809 arabidopsis
29	173	14.7	701	4	O94854	O94854 homo sapien
30	171.5	14.6	438	16	O9A2W5	O9a2w5 caulobacter
31	171.5	14.6	616	4	O9H6K5	O9h6k5 homo sapien
32	171	14.5	544	12	O89392	O89392 paramecium
33	171	14.5	1340	16	Q9L1H8	Q9l1h8 streptomyc
34	170.5	14.5	601	5	O17113	O17113 babesia bov
35	170	14.4	295	2	O32850	O32850 mycobacteri
36	170	14.4	428	10	Q94IQ8	Q94iq8 zea mays (m
37	170	14.4	548	16	O06404	O06404 mycobacteri
38	170	14.4	842	5	O9VGC9	O9vgc9 drosophila
39	170	14.4	864	5	O95U45	O95u45 drosophila
40	170	14.4	864	5	O9VGC8	O9vgc8 drosophila
41	170	14.4	1354	11	Q9EPW8	Q9epw8 mus musculu
42	169.5	14.4	802	11	P70433	P70433 mus musculu
43	169	14.3	312	5	Q9UANI	Q9uani drosophila
44	169	14.3	552	5	O96579	O96579 trypanosoma
45	169	14.3	1099	2	O49560	O49560 mycoplasma

ALIGNMENTS

RESULT 1

Q9D5J3	PRELIMINARY;	PRT;	344 AA.
AC Q9D5J3;			
DT 01-JUN-2001 (TRENBLrel. 17, Created)			
DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)			
DT 01-JUN-2001 (TRENBLrel. 17, Last annotation update)			
DE 4930432K09RIK protein.			
GN 4930432K09RIK.			
OS Mus musculus (Mouse).			
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX NCBI_TaxID=10090;			
RN [1]			
RP SEQUENCE FROM N.A.			
RC STRAIN=C57BL/6J; TISSUE=TESTIS;			
RX MEDLINE=21085660; PubMed=11217851;			
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,			
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,			
RA Saito T., Okazaki Y., Gojohori T., Bono H., Kasukawa T., Saito R.,			
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,			
RA Fieischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,			
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,			
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,			
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,			
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,			
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,			
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,			
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,			
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,			
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,			
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,			
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,			
RA Hayashizaki Y.;			
RT "Functional annotation of a full-length mouse cDNA collection.";			
RL Nature 409:685-690(2001).			
DR EMBL; AK015291; BAB29782.1; "			
DR MGD; MGI:1921029; 4930432K09RIK.			

```
DR InterPro: IPR002965: P_rich_extensn.
DR PRINTS: PR01217; PRICHEXTENS. D65EAD71CE5802AC CRC64;
SQ SEQUENCE 344 AA: 36294 MW;  D65EAD71CE5802AC CRC64;

Query Match
  30.5%; Score 359.5; DB 11; Length 344;
Best Local Similarity 32.0%; Pred. No. 2.8e-17;
Matches 106; Conservative 25; Mismatches 87; Indels 113; Gaps 11;

QY 1 MKLLWACIVCVAFARKRRPFIGEDDN-DDGHPHLPSLNIPYGI--RNLP-----PPLY 53
   |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 1 MKLLWACIMCAFA-KRKSPSTEEDYGNRYPLNSLNIPFGWNLNLPFLLPENT 59
   |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 54 RPVNTVPSYGGNTYDTGLSPYWLTSFGFPYVYHIRGFLATQNVPL----- 104
   :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 60 QQNTITKFGGSELEKGLSLFWTATPSKLYVYQSLNPADAPLNGPPVAPLPKHF 119
   :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 105 PRGFFVPVPSRFF-----SAAAPAAP----- 127
   |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 120 PRRPYFVIPPKISVISPVRRPVAVPAMPAGEGLVPEFPVKPTLGLPQAVKLGTPV 179
   |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 128 -----PIAAEPA-----AAAPLTATPVAAEPAAGAPVAAE 157
   |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 180 PSGPKPLAPEPALSPFGAPAPVQFGAPAPHQLEAAVPOASRLMAPEAPVLSVGAQ 239
   |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 158 PAAEAPVG-----APAAEAPVAAEAPV-----GVEPAAEPPSP----- 195
   |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 240 SLAESAGLSPENKPTSGEAPTQSLPALPAGLAVEAKLPAESAAGRPPELWASQS 299
   |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 196 -----AEPATAKAAPEPHSPSLEQANQ 219
   |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 300 VVGKLITAEPTAKPQVLEPVEAKSASQEAQ 330
   |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:

RESULT 2
Q8V2A4 PRELIMINARY; PRT; 3179 AA.
AC Q8V2A4;
DT 01-MAR-2002 (Tremblrel. 20, Created)
DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE BPLF1.
OS Human herpesvirus 4 (Epstein-Barr virus).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Lymphocryptovirus.
OX NCBI_TaxID=10376;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=M-ABA;
RA Schmaus S., Wolf H., Schwarzmann F.;
RT "The open reading frame BPLF1 of Epstein-Barr virus is expressed in
   late viral replication.";
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF448220; AAL40834.1; -.
DR InterPro: IPR000104; Antifreeze_1.
DR PRINTS: PR00308; ANTIFREEZE1.
SQ SEQUENCE 3179 AA: 340397 MW;  AB0A5D9088A945B1 CRC64;

Query Match
  19.1%; Score 224.5; DB 12; Length 3179;
Best Local Similarity 31.1%; Pred. No. 3.8e-07;
Matches 68; Conservative 23; Mismatches 61; Indels 67; Gaps 9;

QY 24 GEDNDGDGPHLPSLNIPYGINLPPPLYRPVNTVPSYPGNTYTDGLSPYWLITSFG 83
   |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 278 GEAAGTGADSSPPVMLPFERRIXPNL--RPL-----PSRSFTS---DSFPAARYSP- 325
   |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 84 FPYVYHIRGFLATQNLNVPLPPRGFFVPPSPRFFSAAAPA-APPIAAEPAAPLTA 142
   :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 326 -----AKTNSPPSSP-----ASAAPASAPASAPASAPASAA 359
   :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 143 PVAAEPAAGAPVAAEAPVGAEPAAEAPVAAEAPVAAEAPVAAEAPVAAEAPVAAE 193
   :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 360 PASAAPASAPASAPASAPASAPASAPASAPASAPASAPASAPASAPASAP 419
   :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:

QY 194 -SPAEPATAKP-----AAPE-----PHSPSPS 213
   :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 420 HTPGVPAPEPPPRASGGAAPQTPKRKKGLGKDSPHKKPT 458
   :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:

RESULT 3
Q9X6J9 PRELIMINARY; PRT; 1100 AA.
AC Q9X6J9;
DT 01-NOV-1999 (Tremblrel. 12, Created)
DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE DNA primase.
GN PRI.
OS Klebsiella oxytoca.
OG Plasmid PACM1.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Klebsiella.
OX NCBI_TaxID=571;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20334746; PubMed=10873523;
RA Preston K.E., Radonski C.C.A., Venezia R.A.;
RT "Nucleotide sequence of a 7-kb fragment of pacM1 encoding an IncM DNA
   primase and other putative proteins associated with conjugation.";
RL Plasmid 44:12-23(2000).
DR EMBL: AF139719; AAD33802.1; -.
KW Plasmid.
SQ SEQUENCE 1100 AA; 120370 MW;  36EAD2BD33FCB732 CRC64;

Query Match
  18.8%; Score 221; DB 2; Length 1100;
Best Local Similarity 31.4%; Pred. No. 2.3e-07;
Matches 76; Conservative 19; Mismatches 85; Indels 62; Gaps 9;

QY 9 IVCVAFARKRRFP-----FIGEDD--NDD-----GHPHLPSLNIPYGIN 46
   :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 238 MVEVAGRLKDTFPNSTHYFLADNDIYKENVGLEKATEAAELTAGHLVLPASNP----- 352
   :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 47 LPPPLYRPVNTVPSYPGNTYTDGLSPYWLITSFGFPYV-YHIRGFLATQNLNVPLP 105
   |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 353 -----KEGLTDYNDLHVSEGLEQVRLQVEG--AINOMNRVDTM 388
   :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 106 PRGFFVPVPSRFFSA-AAAPAP-----PIAAEPAAPLTAAPVAAEPAAGAPVAAEPA 160
   :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 389 PTDNPNITDVNHSSTDSAAVAAPEKAAPVASTPAAAEVPEAPVAPASAPAAAEPEETAPVA 448
   :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 161 EAPVGAE-----PAAEAPVAAEAPVGAEPVGAEPSPAEPAEPATAKPAEPHSPSPS 215
   :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 449 SAPAAELVETAPVASTPEAAPEVPAAPVAPASAPAAAEPEEAPVAPASAPAAAEPEETAPVA 508
   :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 216 QA 217
   |
Db 509 SA 510

RESULT 4
Q9LIE8 PRELIMINARY; PRT; 1480 AA.
AC Q9LIE8;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Similarity to cell wall-plasma membrane linker protein.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RA Kaneko T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
```

```
DR InterPro: IPR002965: P_rich_extensn.
DR PRINTS: PR01217; PRICHEXTENS. D65EAD71CE5802AC CRC64;
SQ SEQUENCE 344 AA: 36294 MW;  D65EAD71CE5802AC CRC64;

Query Match
  30.5%; Score 359.5; DB 11; Length 344;
Best Local Similarity 32.0%; Pred. No. 2.8e-17;
Matches 106; Conservative 25; Mismatches 87; Indels 113; Gaps 11;

QY 1 MKLLWACIVCVAFARKRRPFIGEDDN-DDGHPHLPSLNIPYGI--RNLP-----PPLY 53
   |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 1 MKLLWACIMCAFA-KRKSPSTEEDYGNRYPLNSLNIPFGWNLNLPFLLPENT 59
   |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 54 RPVNTVPSYGGNTYDTGLSPYWLTSFGFPYVYHIRGFLATQNVPL----- 104
   :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 60 QQNTITKFGGSELEKGLSLFWTATPSKLYVYQSLNPADAPLNGPPVAPLPKHF 119
   :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 105 PRGFFVPVPSRFF-----SAAAPAAP----- 127
   |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 120 PRRPYFVIPPKISVISPVRRPVAVPAMPAGEGLVPEFPVKPTLGLPQAVKLGTPV 179
   |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 128 -----PIAAEPA-----AAAPLTATPVAAEPAAGAPVAAE 157
   |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 180 PSGPKPLAPEPALSPFGAPAPVQFGAPAPHQLEAAVPOASRLMAPEAPVLSVGAQ 239
   |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 158 PAAEAPVG-----APAAEAPVAAEAPV-----GVEPAAEPPSP----- 195
   |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 240 SLAESAGLSPENKPTSGEAPTQSLPALPAGLAVEAKLPAESAAGRPPELWASQS 299
   |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 196 -----AEPATAKAAPEPHSPSLEQANQ 219
   |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 300 VVGKLITAEPTAKPQVLEPVEAKSASQEAQ 330
   |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:

RESULT 2
Q8V2A4 PRELIMINARY; PRT; 3179 AA.
AC Q8V2A4;
DT 01-MAR-2002 (Tremblrel. 20, Created)
DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE BPLF1.
OS Human herpesvirus 4 (Epstein-Barr virus).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Lymphocryptovirus.
OX NCBI_TaxID=10376;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=M-ABA;
RA Schmaus S., Wolf H., Schwarzmann F.;
RT "The open reading frame BPLF1 of Epstein-Barr virus is expressed in
   late viral replication.";
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF448220; AAL40834.1; -.
DR InterPro: IPR000104; Antifreeze_1.
DR PRINTS: PR00308; ANTIFREEZE1.
SQ SEQUENCE 3179 AA: 340397 MW;  AB0A5D9088A945B1 CRC64;

Query Match
  19.1%; Score 224.5; DB 12; Length 3179;
Best Local Similarity 31.1%; Pred. No. 3.8e-07;
Matches 68; Conservative 23; Mismatches 61; Indels 67; Gaps 9;

QY 24 GEDNDGDGPHLPSLNIPYGINLPPPLYRPVNTVPSYPGNTYTDGLSPYWLITSFG 83
   |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 278 GEAAGTGADSSPPVMLPFERRIXPNL--RPL-----PSRSFTS---DSFPAARYSP- 325
   |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 84 FPYVYHIRGFLATQNLNVPLPPRGFFVPPSPRFFSAAAPA-APPIAAEPAAPLTA 142
   :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 326 -----AKTNSPPSSP-----ASAAPASAPASAPASAPASAA 359
   :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 143 PVAAEPAAGAPVAAEAPVGAEPAAEAPVAAEAPVAAEAPVAAEAPVAAEAPVAAE 193
   :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 360 PASAAPASAPASAPASAPASAPASAPASAPASAPASAPASAPASAPASAP 419
   :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:

QY 194 -SPAEPATAKP-----AAPE-----PHSPSPS 213
   :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 420 HTPGVPAPEPPPRASGGAAPQTPKRKKGLGKDSPHKKPT 458
   :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:

RESULT 3
Q9X6J9 PRELIMINARY; PRT; 1100 AA.
AC Q9X6J9;
DT 01-NOV-1999 (Tremblrel. 12, Created)
DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE DNA primase.
GN PRI.
OS Klebsiella oxytoca.
OG Plasmid PACM1.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Klebsiella.
OX NCBI_TaxID=571;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20334746; PubMed=10873523;
RA Preston K.E., Radonski C.C.A., Venezia R.A.;
RT "Nucleotide sequence of a 7-kb fragment of pacM1 encoding an IncM DNA
   primase and other putative proteins associated with conjugation.";
RL Plasmid 44:12-23(2000).
DR EMBL: AF139719; AAD33802.1; -.
KW Plasmid.
SQ SEQUENCE 1100 AA; 120370 MW;  36EAD2BD33FCB732 CRC64;

Query Match
  18.8%; Score 221; DB 2; Length 1100;
Best Local Similarity 31.4%; Pred. No. 2.3e-07;
Matches 76; Conservative 19; Mismatches 85; Indels 62; Gaps 9;

QY 9 IVCVAFARKRRFP-----FIGEDD--NDD-----GHPHLPSLNIPYGIN 46
   :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 238 MVEVAGRLKDTFPNSTHYFLADNDIYKENVGLEKATEAAELTAGHLVLPASNP----- 352
   :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 47 LPPPLYRPVNTVPSYPGNTYTDGLSPYWLITSFGFPYV-YHIRGFLATQNLNVPLP 105
   |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 353 -----KEGLTDYNDLHVSEGLEQVRLQVEG--AINOMNRVDTM 388
   :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 106 PRGFFVPVPSRFFSA-AAAPAP-----PIAAEPAAPLTAAPVAAEPAAGAPVAAEPA 160
   :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 389 PTDNPNITDVNHSSTDSAAVAAPEKAAPVASTPAAAEVPEAPVAPASAPAAAEPEETAPVA 448
   :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 161 EAPVGAE-----PAAEAPVAAEAPVGAEPVGAEPSPAEPAEPATAKPAEPHSPSPS 215
   :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 449 SAPAAELVETAPVASTPEAAPEVPAAPVAPASAPAAAEPEEAPVAPASAPAAAEPEETAPVA 508
   :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 216 QA 217
   |
Db 509 SA 510

RESULT 4
Q9LIE8 PRELIMINARY; PRT; 1480 AA.
AC Q9LIE8;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Similarity to cell wall-plasma membrane linker protein.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RA Kaneko T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
```



```

[2]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RX MEDLINE=20363099; PubMed=10907853;
RA Nakamura Y.;
RT "Structural analysis of Arabidopsis thaliana chromosome 3. II.
RT Sequence features of the regions of 4,251,695 bp covered by ninety Pl,
RT TAC and BAC clones."
RL DNA Res. 7:217-221(2000).
DR EMBL; AP001306; BAB03062.1; -.
DR HSSP; P24337; 1HYP.
DR InterPro; IPR002965; P-rich_extensin.
DR InterPro; IPR001768; Try/amyl_inhbr.
DR Pfam; PF00234; try_alpha_aml; 1.
DR PRINTS; PR01217; PRICHEXTENS.
DR PR01217; PRICHEXTENS.
SQ SEQUENCE 1480 AA; 147154 MW; D1AC0C79F155E732 CRC64;

Query Match 18.2%; Score 214.5; DB 10; Length 1480;
Best Local Similarity 32.6%; Pred. No. 8.6e-07;
Matches 63; Conservative 13; Mismatches 70; Indels 47; Gaps

QY 33 PLHPSLNIPGIRNLP---PPLYRVRNVTPSYPGNTYDTCGLPSYPHILTSPGFPVYH 89
DB 665 PAKPPVAIP-----PIATPPVAKPPVATPP-----TATPP-----IATP----- 698
QY 90 IRGFPPLATQNLNVPL--PPRGPFV-----PPSRFTSAAAPAA--PPIAAEAAAAAPL 139
DB 699 ----PIAT-----PPVTPPTATPPVATPPPIAKPPTTIPPTATPPVAMPPIATPPTAKPPI 750
QY 140 TATPVAEPAAGAPVAEPAEAAEPVGAEPAAEAPVAEPAEAPVGAEPFAEEPSAEP 199
DB 751 ATPPIATPPVAKPPVTPPTATPPIATPPIAKSPVATPPATPPVATPPPIAKPPVATPPT 810
QY 200 TAKPAAPERHPSP 212
DB 811 TAPPTATPPVAKP 823

RESULT 5
Q8U2I9 ID Q8U2I9 PRELIMINARY; PRT; 3105 AA.
AC Q8U2I9
AD Q8U2I9
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE BLF1.
OS cercopithecine herpesvirus 15.
OC viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Lymphocryptovirus.
OX NCBI_TaxID=104228;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LCL8664;
RX MEDLINE=97048062; PubMed=8892903;
RA Franken M., Devergne O., Rosenzweig M., Annis B., Kieff E., Wang F.;
RT "Comparative analysis identifies conserved tumor necrosis factor
RT receptor-associated factor 3 binding sites in the human and simian
RT Epstein-Barr virus oncogene LMP1."
RL J. Virol. 70:7819-7826(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=LCL8664;
RX MEDLINE=99412410; PubMed=10482645;
RA Rivallier P., Quink C., Wang F.;
RT "Strong selective pressure for evolution of an Epstein-Barr virus
RT LMP2B homologue in the rhesus lymphocryptovirus."
RL J. Virol. 73:8867-8872(1999).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=LCL8664;
RX MEDLINE=20304984; PubMed=10846073;
RA Jiang H., Cho Y.G., Wang F.;
RT "Structural, functional, and genetic comparisons of Epstein-Barr virus
RT
```

```

RESULT 6
Q17112
ID Q17112 PRELIMINARY; PRT; 607 AA.
AC Q17112
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DE Surface antigen (Fragment)
DE 80 kDa protein.
OS Babesia bovis.
OC Eukaryota; Alveolata; Apicomplexa; Piroplasmida; Babesiidae; Babesia.
OX NCBI_TaxID=5865;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93341567; PubMed=8341317;
RA Daltymple B.P., Peters J.M., Goodger B.V., Bushell G.R.,
RA Waltsbuhl D.J., Wright I.G.;
RT "Cloning and characterisation of cDNA clones encoding two Babesia
RT bovis proteins with homologous amino- and carboxy-terminal domains.";
RL Mol. Biochem. Parasitol. 59:181-189(1993).
DR EMBL; M93126; AAA02753.1; -.
SQ SEQUENCE 607 AA; 67129 MW; 2E06EC7DFB43D732 CRC64;

Query Match 17.5%; Score 206.5; DB 5; Length 607;
Best Local Similarity 46.5%; Pred. No. 1.3e-06;
Matches 46; Conservative 7; Mismatches 35; Indels 11; Gaps 2;

QY 128 PTAEEPAALATATPVAEPAAGAPVAEPAEAPVAEPAEAPVAEPAEAPVGE 187
DB 329 PIVEEPAEIVEEPIVEEPIVEEPAEPAEPAEPAEPAEPAEPAEPAEPAE 388

QY 188 PAAEEP-----SPAEPATAKPA---APEPHSPSLE 215
DB 389 PVAEPPVAEPPVCEPPEPAPAEKPAETPAETPAETPAE 427

RESULT 7
Q26892
ID Q26892 PRELIMINARY; PRT; 391 AA.
AC Q26892;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DE Surface antigen (Fragment).
OS Trypanosoma cruzi.
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
OX NCBI_TaxID=5693;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=CAI;
RX MEDLINE=92389985; PubMed=1518528;
RA Buschiazzo A., Campetella O.E.;
RT "Sequence of the gene for a Trypanosoma cruzi protein antigenic during
RT the chronic phase of human Chagas disease.";
RL Mol. Biochem. Parasitol. 54:125-128(1992).
DR EMBL; M92046; AAB97873.1; -.
FT NON_TER 1
SQ SEQUENCE 391 AA; 39459 MW; F234844B1F865BC7 CRC64;

Query Match 16.9%; Score 199; DB 5; Length 391;
Best Local Similarity 43.8%; Pred. No. 2.6e-06;
Matches 42; Conservative 14; Mismatches 40; Indels 0; Gaps 0;

QY 113 PPSREFFSAAPPIAAEPAALATATPVAEPAAGAPVAEPAEAPVGEPAEAE 172
DB 5 PKSAEPKSAEPKSAEPKSAEPKSAEPKSAEPKSAEPKSAEPKSAEPKSAEPK 64

QY 173 PVAEPAEAPVGEPAEAPVGEPAEAPVGEPAEAPVGEPAEAPVGEPAEAP 208
DB 65 PKSAEPKSAEPKSAEPKSAEPKSAEPKSAEPKSAEPKSAEPKSAEPKSAEP 100

RESULT 8

```

```

Q26893
ID Q26893 PRELIMINARY; PRT; 722 AA.
AC Q26893;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Surface antigen (Fragment).
OS Trypanosoma cruzi.
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
OX NCBI_TaxID=5693;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=CAI;
RX MEDLINE=92389985; PubMed=1518528;
RA Buschiazzo A., Campetella O.E.;
RT "Sequence of the gene for a Trypanosoma cruzi protein antigenic during
RT the chronic phase of human Chagas disease.";
RL Mol. Biochem. Parasitol. 54:125-128(1992).
DR EMBL; M92047; AAB97874.1; -.
FT NON_TER 1
SQ SEQUENCE 722 AA; 74874 MW; C572189CD6DB4B3E CRC64;

Query Match 16.5%; Score 194.5; DB 5; Length 722;
Best Local Similarity 29.8%; Pred. No. 9.8e-06;
Matches 59; Conservative 26; Mismatches 74; Indels 39; Gaps 6;

QY 20 FPFCEDDNDGHLPHSLNIPYGIIRNLPPPLYRYPVNTVP-----SYPGNTYTD 70
DB 189 FYFGAYDEQLSSRKIHATLANVF-----LYNRPLNDAEIGALNANKVSLP---FTEK 237

QY 71 GLPSYWPILSPGFYVYHIRGFPPLATQNLVPPLPGRGFPVPPRFFSAAAPAAPIA 130
DB 238 -KPAKASTATSPSVHEV-----TTPVATE-----PKSAEPKSAVPKSAEPKS 278

QY 131 AEPAAAPLTATPVAEPAAGAPVAEPAEAPVGAEPAAEAPVGAEPAAEAPVGEPA 190
DB 279 AVPKSAEPKSAEPKSAEPKSAVPKSAEPKSAVPKSAEPKSAEPKSAEPKSAEPKS 338

QY 191 EEPSPAEPATAKPAEPEP 208
DB 339 AEPKSAEPKSAVPKSAEP 356

RESULT 9
Q39266
ID Q39266 PRELIMINARY; PRT; 3534 AA.
AC Q39266;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Counterpart of HSV-1 gene UL36 and VZV gene 22.
GN 24.
OS Equine herpesvirus 4.
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Varicellovirus.
OX NCBI_TaxID=10331;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=NS80567;
RX MEDLINE=98264497; PubMed=9603335;
RA Telford E.A., Watson M.S., Perry J., Cullinane A.A., Davison A.J.;
RT "The DNA sequence of equine herpesvirus-4.";
RL J. Gen. Virol. 79:1197-1203(1998).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=NS80567;
RA Telford E.A., Watson M.S., Perry J., Cullinane A.A., Davison A.J.;
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF030027; AAC59539.1; -.
DR InterPro; IPR005210; Herpes_UL36.
DR Pfam; PF03586; Herpes_UL36; 1.
SQ SEQUENCE 3534 AA; 379956 MW; AF5164843CA1BFC6 CRC64;

```



```
RESULT 15
Q9FPQ5
ID Q9FPQ5 PRELIMINARY; PRT; 386 AA.
AC Q9FPQ5;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DE 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Gamete-specific hydroxyproline-rich glycoprotein a2.
GN A2.
OS Chlamydomonas reinhardtii.
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC Chlamydomonadaceae; Chlamydomonas.
OX NCBI_TaxID=3055;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CC-621;
RA Ferris P.J., Woessner J.P., Waffenschmidt S., Kilz S., Drees J.,
RA Goodenough U.W.;
RT "Glycosylated polyproline II rods-with-kinks as a structural motif in
RT plant hydroxyproline-rich glycoproteins.";
RL Biochemistry 0:0-0(2001).
DR EMBL; AF309495; AAG45421.1; -.
DR InterPro; IPR002965; P-rich_extensn.
DR PRINTS; PR01217; PRICHEXTENSN.
SQ SEQUENCE 386 AA; 38775 MW; 7B4F4CA7D95C2EF7 CRC64;

Query Match 15.4%; Score 182; DB 10; Length 386;
Best Local Similarity 30.3%; Pred. No. 3.7e-05;
Matches 57; Conservative 18; Mismatches 73; Indels 40; Gaps 7;

QY 44 IRLPPLLYRVNTVPSPGNTYDTGLPSYPWILTSPGPPYVYVHIRGFPLATQLNVPP 103
DB 126 LKSLSP---SPMPKPSPSPSPPPSPMPSP-APPSAPP-----SP 165

QY 104 LPPRGFP--FVPPSRFFSAAAAPA-APPIAAEPAAAPLTATPVAAEPAGAPVAAEFPA 160
DB 166 LPPSPVPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSP 225

QY 161 EAPVGAEPAAEPVAAEPAAEPAAEPVAAEPAAEPAAEPAAEPAAEPAAEPAAEP 204
DB 226 EPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSP 284

QY 205 APEPHPS 212
DB 285 PPPSPSP 292
```

Search completed: February 12, 2003, 13:47:41
Job time : 66 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 10, 2003, 11:44:43 ; Search time 1724.77 Seconds
(without alignments)
11085.832 Million cell updates/sec

Title: US-09-923-236-3
Perfect score: 657
Sequence: 1 atgaarytnytnytgtggc.....snyngarcargcnaaycar 657

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:

- 1: gb_ba.*
- 2: gb_htg.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vi.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_om.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pl.*
- 26: em_ro.*
- 27: em_sts.*
- 28: em_un.*
- 29: em_vi.*
- 30: em_htg_hum.*
- 31: em_htg_inv.*
- 32: em_htg_other.*
- 33: em_htg_mus.*
- 34: em_htg_pln.*
- 35: em_htg_rod.*
- 36: em_htg_mam.*
- 37: em_htg_vrt.*
- 38: em_sy.*
- 39: em_htgo_hum.*
- 40: em_htgo_mus.*
- 41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	450	70.0	998	6	AX061225	AX061225 Sequence
2	402	61.2	96481	9	AC104811	AC104811 Homo sapi
3	400.8	61.0	157836	2	AC060328	AC060328 Homo sapi
C 4	370.6	56.4	185969	2	AC069037	AC069037 Homo sapi
5	292	44.4	185969	2	AC069037	AC069037 Homo sapi
6	117	17.8	228404	2	AC095672	AC095672 Rattus no
C 7	107.8	16.4	177035	2	AC099443	AC099443 Rattus no
8	105.4	16.0	325808	2	AC121787	AC121787 Mus muscu
9	105.2	16.0	9618	6	AX460946	AX460946 Sequence
10	104.2	15.9	2286	6	AX460942	AX460942 Sequence
11	103.8	15.8	5120	6	AX067148	AX067148 Sequence
12	103.8	15.8	5120	9	AF152102	AF152102 Homo sapi
C 13	100.4	15.3	1442	9	AF087653	AF087653 Homo sapi
C 14	99.6	15.2	3489	6	AX107938	AX107938 Sequence
C 15	99.6	15.2	3489	6	AX107938	AX107938 Sequence
C 16	99.6	15.2	3489	14	KSU52064	US2064 kaposi's sa
17	99.6	15.2	32207	6	AR065852	AR065852 Sequence
18	99.6	15.2	32207	6	ARI27850	ARI27850 Sequence
19	99.6	15.2	32207	6	ARI94752	ARI94752 Sequence
20	99.6	15.2	137508	14	KSU75698	U75698 kaposi's sa
21	99	15.1	3941	10	RNLAR2	X83546 R.norvegicu
C 22	98.8	15.0	204259	2	AC110817	AC110817 Mus muscu
23	98.6	15.0	28559	14	AF148805	AF148805 Kaposi's
24	98.2	14.9	133661	14	U93872	U93872 Kaposi's sa
C 25	97.6	14.9	164328	2	AC114372	AC114372 Rattus no
26	97.6	14.9	182806	2	AC098902	AC098902 Rattus no
27	97.4	14.8	2479	10	RNU87960	U87960 Rattus norv
C 28	97.4	14.8	201470	2	AC113649	AC113649 Rattus no
29	97.2	14.8	2183	9	AK092426	AK092426 Homo sapi
30	97.2	14.8	163869	9	AC090645	AC090645 Homo sapi
31	97	14.8	161987	9	AC006070	AC006070 Homo sapi
32	96.8	14.7	131274	2	AC096869	AC096869 Rattus no
C 33	96.6	14.7	128040	2	AC097797	AC097797 Rattus no
34	96.6	14.7	164328	2	AC114372	AC114372 Rattus no
C 35	96.4	14.7	168778	9	AC090886	AC090886 Homo sapi
C 36	96.4	14.7	169590	9	AC090004	AC090004 Homo sapi
37	96.2	14.6	207423	2	AC109662	AC109662 Rattus no
38	96.2	14.6	347572	2	AC107303	AC107303 Homo sapi
39	95.8	14.6	176822	2	AC099174	AC099174 Rattus no
40	95.2	14.5	177035	2	AC099443	AC099443 Rattus no
41	95	14.5	1159	9	AF126748	AF126748 Homo sapi
C 42	95	14.5	1472	9	AF126749	AF126749 Homo sapi
43	95	14.5	192027	2	AC067715	AC067715 Mus muscu
C 44	95	14.5	206602	10	AL590997	AL590997 Mouse DNA
45	94.8	14.4	2285	3	BB080KDC1B	M93126 Babesia bov

ALIGNMENTS

RESULT 1
AX061225
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS

AX061225
Sequence 72 from Patent WO0078953.
AX061225
AX061225.1 GI:12406361
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 998)
Lal,P., Tang,J., Yue,H., Hillman,J.L., Tang,Y.T., Bandman,O.,
Burford,N., Baughn,M.R., Azimzai,Y., Lu,D.A., Au-Young,J. and
Patterson,C.

AX061225
Sequence 72 from Patent WO0078953.
AX061225
AX061225.1 GI:12406361
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 998)
Lal,P., Tang,J., Yue,H., Hillman,J.L., Tang,Y.T., Bandman,O.,
Burford,N., Baughn,M.R., Azimzai,Y., Lu,D.A., Au-Young,J. and
Patterson,C.

RESULT 2	
AC104811	
LOCUS	96481 bp DNA linear PRI 16-APR-2002
DEFINITION	Homo sapiens BAC clone RP11-6531.5 from 4, complete sequence.
ACCESSION	AC104811
VERSION	AC104811.4 GI:19698738
KEYWORDS	HTGS.
SOURCE	Homo sapiens.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS	1 (bases 1 to 96481) Sulston,J.E. and Waterston,R.
TITLE	Toward a complete human genome sequence

Db	36323	AGCCCTGCTCCCGAGAACCTCACCTCTCCCTCTCTTGAACAGG	36368
RESULT	4		
AC069037/C			
LOCUS	AC069037	185969 bp	DNA linear HTG 17-MAY-2000
DEFINITION	Homo sapiens chromosome 4 clone RP11-653L5 map 4, WORKING DRAFT SEQUENCE, 34 unordered pieces.		
ACCESSION	AC069037		
VERSION	AC069037.1	GI:7884885	
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT.		
SOURCE	Homo sapiens.		
ORGANISM	Homo sapiens		
REFERENCE	Fukuyota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 185969)		
AUTHORS	Birren,B., Linton,L., Nusbaum,C. and Lander,E.		
TITLE	Homo sapiens chromosome 4, clone RP11-653L5		
JOURNAL	Unpublished		
REFERENCE	2 (bases 1 to 185969)		
AUTHORS	Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N., Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F., Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P., DeArrellano,K., Dewar,K., Diaz,J.S., Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N., Grant,G., Hagos,B., Hearford,A., Horton,L., Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczyk,J., Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N., Pisanic,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and Zody,M.		
Direct Submission			
TITLE	Submitted (17-MAY-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA		
JOURNAL	All repeats were identified using RepeatMasker:		
COMMENT	Smit, A.F.A. & Green, P. (1996-1997) http://ftp.genome.washington.edu/RM/RepeatMasker.html		
	----- Genome Center		
	Center: Whitehead Institute/ MIT Center for Genome Research		
	Center code: WIBR		
	Web site: http://www-seq.wi.mit.edu		
	Contact: sequence_submissions@genome.wi.mit.edu		
	----- Project Information		
	Center clone name: L7416		
	Center project name: 653_L5		
	----- Summary Statistics		
	Sequencing vector: M13; M77815; 100% of reads		
	Chemistry: Dye-terminator Big Dye; 100% of reads		
	Assembly program: Phrap; version 0.960731		
	Consensus quality: 167163 bases at least Q40		
	Consensus quality: 175669 bases at least Q30		
	Consensus quality: 179699 bases at least Q20		
	Insert size: 188000; agarose-fp		
	Insert quality: 182669; sum-of-contigs		
	Quality coverage: 3.6 in Q20 bases; agarose-fp		
	Quality coverage: 3.7 in Q20 bases; sum-of-contigs		

	* NOTE: This is a 'working draft' sequence. It currently		
	* consists of 34 contigs. The true order of the pieces		
	* is not known and their order in this sequence record is		
	* arbitrary. Gaps between the contigs are represented as		
	* runs of N, but the exact sizes of the gaps are unknown.		

Db 36323	AGCCCTGCTGCCAGAACCTCACCCTTCTCCCTCTCTTGAACAGG	36368	
RESULT 4			
AC069037/c			
LOCUS			
DEFINITION			
SEQUENCE, 34 unordered pieces.			
AC069037	185969 bp	DNA	linear HTG 17-MAY-2000
AC069037.1	GI:7884885		
HTG: HTGS_PHASE1; HTGS_DRAFT.			
Homo sapiens.			
Homo sapiens			
Homo sapiens			
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,			
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,			
Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G.,			
Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,			
Collimore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,			
Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,			
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,			
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,			
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,			
Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczyk,J.,			
Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,			
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,			
McLirim,J., Meneus,L., Mihova,T., Miranda,C., Mlangay,V., Morrow,J.,			
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,			
O'Neill,D., Olivier,T.M., Oliver,J., Peterson,K., Pierre,N.,			
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Spencer,B.,			
Roy,A., Santos,R., Schauer,S., Severy,P., Subramanian,A., Talamas,J.,			
Stange-Thomann,N., Stojanovic,N., Travers,M., Trigilio,J.,			
Tesfaye,S., Theodore,J., Tirrell,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,			
Vassiliev,H., Viel,R., Vo,A., Zaimoun,J., Zimmer,A. and Zody,M.			
Young,G., Zaimoun,J., Zimmer,A. and Zody,M.			
Direct Submission			
Submitted (17-MAY-2000) Whitehead Institute/MIT Center for Genome			
Research, 320 Charles Street, Cambridge, MA 02141, USA			
All repeats were identified using RepeatMasker:			
Smit, A.F.A. & Green, P. (1996-1997)			
http://ftp.genome.washington.edu/RM/RepeatMasker.html			
----- Genome Center			
Center: Whitehead Institute/ MIT Center for Genome Research			
Center code: WIBR			
Web site: http://www-seq.wi.mit.edu			
Contact: sequence_submissions@genome.wi.mit.edu			
----- Project Information			
Center project name: L7416			
Center clone name: 653_L5			
----- Summary Statistics			
Sequencing vector: M13; M7815; 100% of reads			
Chemistry: Dye-terminator Big Dye; 100% of reads			
Assembly program: Phrap; version 0.960731			
Consensus quality: 167163 bases at least Q40			
Consensus quality: 175669 bases at least Q30			
Consensus quality: 179699 bases at least Q20			
Insert size: 188000; agarose-fp			
Insert size: 182669; sum-of-contigs			
Quality coverage: 3.6 in Q20 bases; agarose-fp			
Quality coverage: 3.7 in Q20 bases; sum-of-contigs			

* NOTE: This is a 'working draft' sequence. It currently			
* consists of 34 contigs. The true order of the pieces			
* is not known and their order in this sequence record is			
* arbitrary. Gaps between the contigs are represented as			
* runs of N, but the exact sizes of the gaps are unknown.			

Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
 Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
 Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
 Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
 Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
 Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
 Earnhart,C., Edgar,D., Edwards,C., Elhaj,C., Escotto,M.,
 Falls,T., Ferraguto,D., Flagnd,N., Ford,J., Foster,P., Frantz,P.,
 Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
 Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
 Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
 Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
 Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
 Jacobson,B., Jia,Y., Johnson,R., Jollivet,S., Joudah,S.,
 Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
 Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
 Li,J., Li,Z., Lichtarge,O., Lied,C., Lucier,R., Luna,R., Ma,J.,
 Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
 Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
 Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
 Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
 Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
 Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
 Oragunye,N., Oviedo,R., Face,A., Payton,B., Peery,J., Perez,L.,
 Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
 Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savary,G.,
 Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
 Sodergren,E., Sonaik,T., Sparks,A., Stanley,H., Stone,H.,
 Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
 Tansey,J., Taylor,C., Taylor,T., Teifrod,B., Thomas,N., Thomas,S.,
 Usman,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R., Wang,Q.,
 Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
 Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
 Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
 Weinstock,G. and Gibbs,R.
 Direct Submission
 Unpublished
 2 (bases 1 to 228404)
 Worley,K.C.
 Direct Submission
 Submitted (17-SEP-2001) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 228404)
 Worley,K.C.
 Direct Submission
 Submitted (10-JUL-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Jul 10, 2002 this sequence version replaced gi:17942240.
 ----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GC21
 Center clone name: CH230-8P2
 ----- Summary Statistics
 Sequencing vector: M13
 Chemistry: Dye-terminator Big Dye: 100% of reads
 Assembly program: Phrap: version 0.990329
 Consensus quality: 20264 bases at least Q40
 Consensus quality: 206598 bases at least Q30
 Consensus quality: 209968 bases at least Q20

 * NOTE: Estimated insert size may differ from sequence length
 (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 consists of 32 contigs. The true order of the pieces
 is not known and their order in this sequence record is
 arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 *
 * 1347: contig of 1347 bp in length
 * 1348 1447: gap of unknown length
 * 1448 2464: contig of 1017 bp in length
 * 2465 2564: gap of unknown length
 * 2565 3288: contig of 1364 bp in length
 * 3289 4028: gap of unknown length
 * 4029 5397: contig of 1369 bp in length
 * 5398 5497: gap of unknown length
 * 5498 7339: contig of 1842 bp in length
 * 7340 7439: gap of unknown length
 * 7440 9865: contig of 2426 bp in length
 * 9866 9965: gap of unknown length
 * 9966 12984: contig of 3019 bp in length
 * 12985 13084: gap of unknown length
 * 13085 14812: contig of 1728 bp in length
 * 14813 14912: gap of unknown length
 * 14913 17933: contig of 2381 bp in length
 * 17494 17933: gap of unknown length
 * 20553 20753: gap of unknown length
 * 20754 23537: contig of 2784 bp in length
 * 23538 23637: gap of unknown length
 * 23638 27758: contig of 4121 bp in length
 * 27759 27858: gap of unknown length
 * 32984 32984: contig of 5126 bp in length
 * 32985 33084: gap of unknown length
 * 33085 39505: contig of 6421 bp in length
 * 39506 39506: gap of unknown length
 * 39506 45106: contig of 5501 bp in length
 * 45107 45206: gap of unknown length
 * 45207 49670: contig of 4464 bp in length
 * 49671 49770: gap of unknown length
 * 49771 57338: contig of 7568 bp in length
 * 57339 57338: gap of unknown length
 * 61448 61448: contig of 4010 bp in length
 * 61449 63392: contig of 7844 bp in length
 * 63393 69492: gap of unknown length
 * 69493 74542: contig of 5050 bp in length
 * 74543 74642: gap of unknown length
 * 74643 79887: contig of 5245 bp in length
 * 79888 79887: gap of unknown length
 * 79888 85761: contig of 5774 bp in length
 * 85762 85861: gap of unknown length
 * 85862 94213: contig of 8352 bp in length
 * 94214 94313: gap of unknown length
 * 94314 101564: contig of 7251 bp in length
 * 101565 101664: gap of unknown length
 * 101665 113492: contig of 11828 bp in length
 * 113493 113592: gap of unknown length
 * 113593 127974: contig of 14382 bp in length
 * 127975 128074: gap of unknown length
 * 128075 141864: contig of 13790 bp in length
 * 141865 141964: gap of unknown length
 * 141965 15315: contig of 13351 bp in length
 * 15316 15416: gap of unknown length
 * 15416 169126: contig of 13711 bp in length
 * 169127 169226: gap of unknown length
 * 169227 182970: contig of 13744 bp in length
 * 182971 183070: gap of unknown length
 * 183071 204404: contig of 21334 bp in length
 * 204405 204504: gap of unknown length
 * 204505 228404: contig of 23900 bp in length.

FEATURES
 source

1. 228404
 /organism="Rattus norvegicus"
 /db_xref="taxon:10116"
 /clone="CH230-8P2"

57156 a 54728 c 55622 g 57657 t 3241 others

BASE COUNT
 ORIGIN

Query Match		17.8%	Score 117:	DB 2:	Length 228404;
Best local similarity		51.18%	Pred. No. 1.2e-07;		
Matches 136; Conservative		13;	Mismatches 117;	Indels	0: Gaps 0;
QY	358	GCNGCNGCNGCNCNGCNCNCNCNATHGCGNGCNGARCCNGCNGCNGCNCNCNVTN	417		
DB	142786	GCCGATCTGCAGCAGGGCTCCAGCCGACAGAGACTCCAGCGCGCAATCTGCA	142845		
QY	418	ACGNCNACNCCNGTNGCNGCNGARCCNGCNGCNGCNGCNGCNGCNGCNGCNGCNCN	477		
DB	142846	ACAGAGGCTCCAGCGCAGCTGGAATCTGCAGCAGAGACTCCAGCCACACGCGAATCTGCA	142905		
QY	478	GCNGARGCNCNGTNGCNGCNGARCCNGCNGCNGCNGCNGCNGCNGCNGCNGCNCN	537		
DB	142906	GCAGAGGCTCCAGCGCAGCGGAATCTGCAGCAGAGGCTCCAGCGCGCAATCTGCA	142965		
QY	538	GCNGARGCNCNGTNGCNGCNGARCCNGCNGCNGCNGCNGCNGCNGCNGCNGCNCN	597		
DB	142966	GCAGAGGCTCCAGCGCTGCCGAATCTGCAGCAGAGGCTCCAGCGCGCAATCTGCA	143025		
QY	598	ACNGCNAACCCNGCNGCNCNCNCNCNCNCNCNCNCNCNCNCNCNCNCNCNCNCNC	623		
DB	143026	GCAGAGGCTCCAGCGCAGCGCAGAGGC	143051		

RESULT 7	AC099443/C	AC099443	177035 bp	DNA	linear	HTG 12-JUL-2002
LOCUS	AC099443/C	Rattus norvegicus clone CH230-197D1, *** SEQUENCING IN PROGRESS				
DEFINITION	AC099443	***, 57 unordered pieces.				
ACCESSION	AC099443	AC099443.3 GI:21729719				
VERSION	AC099443.3	GI:21729719				
KEYWORDS	HTGS_PHASE1.					
SOURCE	Norway rat.					
ORGANISM	Rattus norvegicus					
REFERENCE	1 (bases 1 to 177035)					
AUTHORS	Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayalew,M., Banks,T., Barbara,J., Benton,J., Binage,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowle,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P., Buhaq,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Fallis,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louissegh,H., Lozada,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenwo,S., Ogih,M., Okwunonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,					

Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Tellford, B., Thomas, N., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wang, Q., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczky, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y. F., Zhou, J., Zorrilla, S., Nelson, D., Weinstein, G. and Gibbs, R.

Direct Submission
Unpublished
2 (bases 1 to 177035)
Worley, K.C.

Direct Submission
Submitted (15-NOV-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 177035)
Worley, K.C.

Direct Submission
Submitted (12-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Jul 11, 2002 this sequence version replaced gi:17973694.

----- Genome Center -----
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information -----
Center project name: GIHL
Center clone name: CH230-197D1
----- Summary Statistics -----
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye: 100% of reads
Assembly program: Phrap: version 0.990329
Consensus quality: 125398 bases at least Q40
Consensus quality: 131372 bases at least Q30
Consensus quality: 136509 bases at least Q20

** NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
** NOTE: This is a 'working draft' sequence. It currently
* consists of 57 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 1017: contig of 1017 bp in length
* 1018 1117: gap of unknown length
* 1118 2246: contig of 1129 bp in length
* 2247 2346: gap of unknown length
* 2347 3468: contig of 1122 bp in length
* 3469 3568: gap of unknown length
* 3569 5053: contig of 1485 bp in length
* 5054 5153: gap of unknown length
* 5154 6584: contig of 1431 bp in length
* 6585 8287: contig of 1603 bp in length
* 8288 8387: gap of unknown length
* 8388 9871: contig of 1484 bp in length
* 9872 9917: gap of unknown length
* 9917 11653: contig of 1682 bp in length
* 11654 11754: gap of unknown length
* 11754 13232: contig of 1479 bp in length
* 13232 13332: gap of unknown length
* 13333 15265: contig of 1933 bp in length
* 15266 15365: gap of unknown length
* 15366 16773: contig of 1408 bp in length
* 16774 16973: gap of unknown length
* 16973 18195: contig of 1322 bp in length
* 18196 18295: gap of unknown length
* 18296 19372: contig of 1077 bp in length

[illegible]


```
Db 2097 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2156
QY 416 TNACNGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 475
Db 2157 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2216
QY 476 CNGCNGARGCNCNGTNGCNGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 535
Db 2217 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2276
QY 536 CNGCNGARGCNCNGTNGCNGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 595
Db 2277 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2336
QY 596 CNACNGCNAARCCNGCNCNGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 641
Db 2337 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2382

RESULT 12
AF152102 5120 bp DNA linear PRI 05-JUN-2000
LOCUS Homo sapiens protein phosphatase 2 regulatory subunit B (PPP2R2B)
DEFINITION gene, partial cds.
ACCESSION AF152102
VERSION AF152102.1 GI:8248031
KEYWORDS
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 5120)
AUTHORS Holmes,S.E., O'Hearn,E.E., McGinnis,M.G., Gorelick-Feldman,D.A.,
Kleiderlein,J.J., Callahan,C., Kwak,N.G., Ingersoll-Ashworth,R.G.,
Sherr,M., Sumner,A.J., Sharp,A.H., Ananth,U., Seltzer,W.K.,
Boss,M.A., Viera-Saecker,A.M., Epplen,J.T., Reiss,O., Ross,C.A.
and Margolis,R.L.
TITLE Expansion of a novel CAG trinucleotide repeat in the 5' region of
PPP2R2B is associated with SCA12
MEDLINE Nat. Genet. 23 (4), 391-392 (1999)
PUBMED 20047763
REFERENCE 10581021
2 (bases 1 to 5120)
AUTHORS Holmes,S.E., O'Hearn,E.E., McGinnis,M.G., Gorelick-Feldman,D.A.,
Kleiderlein,J.J., Callahan,C., Kwak,N.G., Ingersoll-Ashworth,R.G.,
Sherr,M., Sumner,A.J., Sharp,A.H., Ananth,U., Seltzer,W.K.,
Boss,M.A., Viera-Saecker,A.M., Epplen,J.T., Reiss,O., Ross,C.A.
and Margolis,R.L.
TITLE Direct Submission
JOURNAL Submitted (17-MAY-1999) Psychiatry, Johns Hopkins University, 600
N. Wolfe St., Meyer 2-181, Baltimore, MD 21287, USA
FEATURES
Source
1..5120
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/map="5q31-33"
/clone="PAC 21933"
/cell_type="lymphocyte"
repeat_region 2088..2366
/rpt_type=tandem
/rpt_unit=cag
<2473..>5120
/gene="PPP2R2B"
<2473..>3093
/gene="PPP2R2B"
/product="protein phosphatase 2 regulatory subunit B"
<2473..>3023
/gene="PPP2R2B"
/note="5' RACE"
3024..>3093
/gene="PPP2R2B"
/note="PP2A-B55beta"
```

```
/codon_start=1
/product="protein phosphatase 2 regulatory subunit B"
/protein_id="AAF74024.1"
/db_xref="GI:8248032"
/translation="MEEDIDTRKINNSFLRDHSYATE"

BASE COUNT 1336 a 1267 c 1268 g 1249 t
ORIGIN

Query Match 15.8%; Score 103.8; DB 9; Length 5120;
Best Local Similarity 44.4%; Pred. No. 1:1e-05;
Matches 127; Conservative 21; Mismatches 138; Indels 0; Gaps 0;

QY 356 SNGCNGCNGCNCNGCNCNGCNCNGCNCNGCNCNGCNCNGCNCNGCNCNGCNCNGCNCNGCNCNY 415
Db 2097 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2156
QY 416 TNACNGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 475
Db 2157 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2216
QY 476 CNGCNGARGCNCNGTNGCNGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 535
Db 2217 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2276
QY 536 CNGCNGARGCNCNGTNGCNGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 595
Db 2277 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2336
QY 596 CNACNGCNAARCCNGCNCNGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 641
Db 2337 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2382

RESULT 13
AF087653/c 1442 bp DNA linear PRI 14-NOV-2000
LOCUS Homo sapiens clone TGC13-7a trinucleotide repeat region, complete
DEFINITION sequence.
ACCESSION AF087653
VERSION AF087653.1 GI:4588013
KEYWORDS
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1442)
AUTHORS Vincent,J.B., Neves-Pereira,M.L., Paterson,A.D., Yamamoto,E.,
Macedo,A., Kovacs,M., Davies,M., Lieberman,J.A., Meltzer,H.Y.,
Petronis,A. and Kennedy,J.L.
TITLE An unstable trinucleotide-repeat region on chromosome 13 implicated
in spinocerebellar ataxia: a common expansion locus
JOURNAL Am. J. Hum. Genet. 66 (3), 819-829 (2000)
MEDLINE 20177485
PUBMED 10712198
REFERENCE 2 (bases 1 to 1442)
AUTHORS Vincent,J.B., Paterson,A.D., Yamamoto,E., Petronis,A. and
Kennedy,J.L.
TITLE Direct Submission
JOURNAL Submitted (27-AUG-1998) Neurogenetics, Clarke Institute of
Psychiatry, 250 College Street, Toronto, ON M5T 1R8, Canada
FEATURES
Source
1..1442
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="13"
/map="13q21.2-q21.31"
/clone="TGC13-7a"
/cell_type="lymphocyte"
repeat_region 449..505
/note="microsatellite"
/rpt_type=tandem
/rpt_unit=tac
```

repeat_region	506..760	/note="microsatellite"
		/rpt_type=tandem
repeat_region	1188..1436	/rpt_unit=tcg
		/note="satellite DNA"
		/rpt_family="aliphoid"
BASE COUNT	346 a 316 c 262 g 518 t	
ORIGIN		
Query Match	15.3%; Score 100.4; DB 9; Length 1442;	
Best Local Similarity	46.2%; Pred. No. 3.4e-05;	
Matches 123; Conservative	17; Mismatches 126; Indels 0; Gaps 0;	
Qy	349 TTTTYYWSNGCGNGCNCGCCGCGCCNCCNATHGCGNGCGARCCNGCGNCNGCN 408	
Db	: : : : : : : :	
772 TTTTAAAAATGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 713	: : : : : : : :	
Qy	409 GCNCCNYTNACGNACNCCNGTNGCNGCGARCCNGCGNGCGNGCGNCCTGTGCGNCN 468	
Db	: : : : : : : :	
712 GCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 653	: : : : : : : :	
Qy	469 GARGCNGCGNGARGCNCNGTNGGNGCGARCCNGCGNGCGNGCGNCCTGTGCGNCN 528	
Db	: : : : : : : :	
652 GCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 593	: : : : : : : :	
Qy	529 GARGCNGCGNGARGCNCNGTNGGNTNGARCCNGCGNGCGNGCGNGARGARCCNWSNCCNGCN 588	
Db	: : : : : : : :	
592 GCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 533	: : : : : : : :	
Qy	589 GARGCNGCGNACNNARCNGCGNC 614	
Db	: : : : : : : :	
532 GCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 507	: : : : : : : :	
RESULT 14		
AR072952/c		
LOCUS	AR072952 3489 bp DNA linear PAT 28-AUG-2000	
DEFINITION	Sequence 1 from patent US 5948676.	
ACCESSION	AR072952	
VERSION	AR072952.1 GI:9999715	
KEYWORDS	Unknown.	
SOURCE	Unknown.	
ORGANISM	Unclassified.	
REFERENCE	1 (bases 1 to 3489)	
AUTHORS	Chang,Y., Bohenzky,R.A., Russo,J.J., Edelman,I.S. and Moore,P.S.	
TITLE	Immediate early protein from Kaposi's sarcoma-associated	
JOURNAL	herpesvirus, DNA encoding same and uses thereof	
FEATURES	Patent: US 5948676-A 1 07-SEP-1999;	
source	Location/Qualifiers	
	1..3489	
BASE COUNT	1053 a 862 c 1137 g 437 t	
ORIGIN		
Query Match	15.2%; Score 99.6; DB 6; Length 3489; .	
Best Local Similarity	48.5%; Pred. No. 4.3e-05;	
Matches 130; Conservative	7; Mismatches 131; Indels 0; Gaps 0;	
Qy	356 SNGCGNGCGNCGCCGCGCCNCCNATHGCGNGCGARCCNGCGNCNGCGNCN 415	
Db	: : : : : : : : :	
1833 CTGGCTGCATCCTGCTGCTGCTCATCCTGCTGCTCATCCTGCTGCTGCTGCCG 1774	: : : : : : : : :	
Qy	416 TNACNGCACNCCGTNCGNGCGARCCNGCGNGCGNCCTGTGCGNGCGNRCNG 475	
Db	: : : : : : : :	
1773 CTGCTGTGGCTCCCGCTGCTGGCTCCCGTGCTGTGGCTCCCGCTGCTGGCTCCCG 1714	: : : : : : : :	
Qy	476 CNGCGNGARGCNCNGTNGGNGCGARCCNGCGNGCGNGCGNGCGNCCTGTGCGN 535	
Db	: : : : : : : :	
1713 CTGCTGTGGCTCCCGCTGCTGGCTCCCGTGCTGGGGGCTCCCGCTGCTGTGGCTCCCG 1654	: : : : : : : :	
Qy	536 CNGCGNGARGCNCNGTNGGNTNGARCCNGCGNGCGNGARGARCCNWSNCCNGCGNRCNG 595	

[illegible]

Search completed: February 10, 2003, 13:29:52
Job time : 2268.77 secs

XX	Polynucleotides encoding salivary proteins useful as anti-microbial agents -
PT	Claim 3; Column 51-52; 29pp; English.
PS	The invention relates to a polynucleotide derived from the 4q12-4q13 region of human chromosome 4 and encoding a zslg63 polypeptide, a secreted salivary protein with anti-microbial activity. Due to their microbial activity, the sequences can be used in the study of microbial infections, e.g. for recombinant production of anti-microbial proteins. The sequences can be used in the treatment of tooth decay, periodontal disease, thrush, gastrointestinal disease, urinary tract infections, vaginal infections, skin infections, epithelial wounds, chronic tissue damage, acquired immunodeficiency syndrome (AIDS), cystic fibrosis, lung infections, sarcoidosis, emphysema and chronic bronchitis. This sequence represents a degenerate cDNA encoding human zslg63.
XX	Sequence 657 BP; 72 A; 150 C; 114 G; 65 T; 256 other;
QY	Query Match 70.0%; Score 460; DB 24; Length 657;
DB	Best Local Similarity 100.0%; Pred. No. 1.6e-68;
DB	Matches 657; Conservative 0; Mismatches 0; Indels 0; Gaps
QY	1 ATGAARVTNTVNTGGCGCTGYATHGTGTGTCGCTTYGCNMGNAARNGMNTTY 60
DB	1 ATGAARVTNTVNTGGCGCTGYATHGTGTGTCGCTTYGCNMGNAARNGMNTTY 60
QY	61 CCNTYATHGCGNGARGAYGAYGAYGAYGGNCCYATCCYTCNCAACNWSNTNAAAYATH 120
DB	61 CCNTYATHGCGNGARGAYGAYGAYGAYGGNCCYATCCYTCNCAACNWSNTNAAAYATH 120
QY	121 CCNTAYCGNATHMGNAAYTTCNCCNCCNTYATYATYMGNCNGTNAAYACNGTNCN 180
DB	121 CCNTAYCGNATHMGNAAYTTCNCCNCCNTYATYATYMGNCNGTNAAYACNGTNCN 180
QY	181 WSNATYCCGNGNAAYACNTAYACNGAYACNGNVTNCCNWSNTAYCCNTGGATHYTACN 240
DB	181 WSNATYCCGNGNAAYACNTAYACNGAYACNGNVTNCCNWSNTAYCCNTGGATHYTACN 240
QY	241 WSNCCNGGNTTYCCNTAYGTNTAYCATYATHMGNGGNTTYCCNTYNGCNACNCAATYNAAY 300
DB	241 WSNCCNGGNTTYCCNTAYGTNTAYCATYATHMGNGGNTTYCCNTYNGCNACNCAATYNAAY 300
QY	301 GTNCCNCCNTTNCNCCNMGNGGNTTYCCNTTGTTCNCCNWSNMGNTTYTYSNGCN 360
DB	301 GTNCCNCCNTTNCNCCNMGNGGNTTYCCNTTGTTCNCCNWSNMGNTTYTYSNGCN 360
QY	361 GCNGCNGCNGCNGCNGCNCNATHGCGNGCNGARCCNGCNGCNGCNCNCCNTTACN 420
DB	361 GCNGCNGCNGCNGCNGCNCNATHGCGNGCNGARCCNGCNGCNGCNCNCCNTTACN 420
QY	421 GCNACNCCNGTNGCNGCNGARCCNGCNGGNGCNCNCCGTNGCNGCNGARCCNGCNGCN 480
DB	421 GCNACNCCGTNGCNGCNGARCCNGCNGGNGCNCNCCGTNGCNGCNGARCCNGCNGCN 480
QY	481 GARGCNCNGTNGCNGCNGARCCNGCNGCNGARCCNGCNGTNGCNGCNGARCCNGCNGCN 540
DB	481 GARGCNCNGTNGCNGCNGARCCNGCNGCNGARCCNGCNGTNGCNGCNGARCCNGCNGCN 540
QY	541 GARGCNCNGTNGGNTGTNGARCCNGCNGCNGARCCNWSNCCNCCNGCNGARCCNGCNGCN 600
DB	541 GARGCNCNGTNGGNTGTNGARCCNGCNGCNGARCCNWSNCCNCCNGCNGARCCNGCNGCN 600
QY	601 GCNAARCCNGCNGCNGCNGARCCNCAATYCCNWSNCCNWSNTYNGARCCNCAATYCAR 657
DB	601 GCNAARCCNGCNGCNGCNGARCCNCAATYCCNWSNCCNWSNTYNGARCCNCAATYCAR 657

RESULT 2
AAF27729
ID AAF27729 standard; cDNA; 998 BP.

”

AC	AAF27729;	
XX		
XX	28-MAR-2001 (first entry)	
DT		
XX		
DE	Human transport protein TPPT-29 coding sequence.	
XX		
XX	Human; transport protein; TPPT; transport disorder; metabolic disorder;	
KW	neurological disorder; cardiovascular disorder; reproductive disorder;	
KW	immune disorder; cancer; ss.	
KW		
XX		
OS	Homo sapiens.	
XX		
PN	WO20007953-A2.	
XX		
PD	28-DEC-2000.	
XX		
PF	16-JUN-2000; 2000WO-US16668.	
XX		
XX	17-JUN-1999; 99US-0139923.	
PR	10-AUG-1999; 99US-0148177.	
PR	18-AUG-1999; 99US-0149357.	
PR	28-OCT-1999; 99US-0162287.	
XX		
XX	(INCY-) INCYTE GENOMICS INC.	
PA		
XX		
PI	Lal P, Yang J, Yue H, Hillman JL, Tang YT, Bandman O, Burford N;	
PI	Baughn MR, Azimzai Y, Lu DAM, Au-Young J, Patterson C;	
XX		
DR	WPI: 2001-041424/05.	
DR	P-PSDB; AAB60109.	
XX		
XX	Isolated polypeptide with a human transport protein sequence is useful	
PT	for the diagnosis, prevention and treatment of disorders associated	
PT	with the immune, reproductive and cardiovascular systems -	
XX		
PS	Claim 5; Page 158; 165pp; English.	
XX		
CC	The present invention provides the protein and coding sequences for 43	
CC	novel human transport proteins (designated TPPTs): these can be used in	
CC	the diagnosis and treatment of transport, metabolic, neurological,	
CC	reproductive, cardiovascular and immune disorders, and cell proliferative	
CC	disorders such as cancer.	
XX		
XX	Sequence 998 BP; 253 A; 249 C; 190 G; 306 T; 0 other;	
SQ		
Query Match 70.0%; Score 460; DB 22; Length 999;		
Best Local Similarity 61.0%; Pred. NO. 1.7e-69;		
Matches 401; Conservative 103; Mismatches 153; Indels 0; Gaps		

Qy	1	ATGAARYNTVNTYNTGGCGNTGYATHGNTGYTGNGCNTTYTGCNMGNNAARMGNMGNTTY	60
Db	123	ATGAAGGCTTCCTCTTTGGGCGCTGCATCTGATGTGTGTTTGGCAAGGAAGACAGCGTTC	182
Qy	61	CCNTTYATHGGNCARGAYGAYAAYGAYGAYGNCAYCCNTYNCAYCCNWSNYTNAAYATH	120
Db	183	CCCTTCATGTTGAGGATGACAATGACGATGGTCACCCACCTTCATCCATCTCTCAATATT	242
Qy	121	CCNTAYGGNATHMGNAAATYTCNCNCNCNTNTNTAYTAYMGNCCNGNTNAYACNGTNCN	180
Db	243	CCNTATGGCATAGGGAATTTACCACTCCTCTTTATTCGCCAGTGAATACAGTCCCC	302
Qy	181	WSNTAYCCNGGNAAYACNTAYACNGAYACGNGNYTNCNWSNTAYCCNTGGATHYTNACN	240
Db	303	AGTTACCCCTGGGAATACTTACACTGCACACAGGGTTACCTTCGTATCCCTGGATTCTAACT	362
Qy	241	WSNCCNGGNTTYCCNTAYGNTNTAYCATYATHMNGNGNTTYCCNYTNGCNACNACRYTNAAY	300
Db	363	TCCTCTGGATTCCTCTGTCATCATACATCCGCTGGTTTTCCCTTAGCTACTACTGGAAT	422
Qy	301	GTNCCNCNTYTCNCNCNMGNGNTTYCCNTYTGTCNCNCNWSNMGNTTYTYSWNGCN	360
Db	423	GTTCCTCTCCTCCTCCTAGGGTTTCCGTTTGTCCCTCTCAAGTGTCTTTTTCAGCA	482


```

FH Key      Location/Qualifiers
FT CDS      127..786
FT /*tag= a
FT /product= "secreted protein clone cpl16 1"
XX
XX WO9901466-Al.
XX
XX 14-JAN-1999.
XX
XX 01-JUL-1998; 98WO-US13813.
XX
XX 27-OCT-1997; 97US-0958304.
XX
XX 02-JUL-1997; 97US-0887195.
XX
XX (GEMV ) GENETICS INST INC.
XX
XX Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Racie LA;
XX Spaulding V, Treacy M;
XX WPI; 1999-105994/09.
XX P-PSDB; AAW30653.
XX
XX New polynucleotides encoding secreted human proteins - are derived
XX from human foetal brain, adult testes, adult brain, foetal kidney,
XX adult salivary gland, or adult blood cDNA libraries, useful as, e.g.
XX potential vaccines
XX
XX Claim 23; Page 70-71; 107pp; English.
XX
XX The present sequence encodes a human secreted protein from clone
XX cpl16 1, deposited as ATCC 98482. Human secreted protein clone
XX polynucleotides and proteins are predicted to have biological
XX activities which would make them suitable for treating, preventing or
XX ameliorating medical conditions in humans and animals. Suggested
XX activities include nutritional activity, cytokine and cell
XX proliferation/differentiation activity, immune stimulating (e.g. as
XX vaccines) or suppressing activity, haematopoiesis regulating activity,
XX tissue growth activity, activin/inhibin activity, chemotactic/
XX chemokinetic activity, haemostatic and thrombolytic activity, receptor/
XX ligand activity, anti-inflammatory activity, cadherin/tumour invasion
XX suppressor activity, and tumour inhibition activity. The polynucleotides
XX are also stated to be useful for gene therapy.
XX
XX Sequence 1325 BP; 407 A; 289 C; 241 G; 388 T; 0 other;
SQ
Query Match      69.5%; Score 456.8; DB 20; Length 1325;
Best Local Similarity 60.7%; Pred. No. 6e-68;
Matches 399; Conservative 103; Mismatches 155; Indels 0; Gaps 0;
QY 1 ATGAARYTNTYNTGGCNGTGYAPHTGNTGYTGNGCMTTYGCMNGMAARMNGMNTTY 60
Db 127 ATCAAGCTCTCTCTTGGCGCTGCATTGTATGTGTGTTTGAAGGAAGAGACGGTTC 186
QY 61 CCNTTATGNGGARGAYGAYAYGAYGAYGNCAYCCNYTNCAYCCNWSNTYNTAYATH 120
Db 187 CCCTTCATTGGTGGAGTACATACAGTGTGTCACCCACTTCATCCATCTCTGAATATT 246
QY 121 CCNTATGNGNATHMGNAAYTNCNCCNCCNTNTATYATYMGNCNTNAYACNGTNCN 180
Db 247 CCNTATGGCATACAGGAATTTACCACCTCTCTTATTATCGGCCAGTGAATACGTCCC 306
QY 181 WSNATCCGNGNAAYACNTAYACNGAYACNGNYTNCNWSNTAYCCNTGGATHYTNACN 240
Db 307 AGTTACCTCGGGAATACTTACACTGACACAGGGTTACCTTCGTATCCCTGGATTCTAACT 366
QY 241 WSNCCNGGNTTCCNTAYGNTAYATYATHMGNGNTTYCCNYTNGCNACNACRYNTNAY 300
Db 367 TCTCCTGGATTCCTCTATCTATCATATCCGTTGGTTTCCCTTAGCTACTCAGTTGAAT 426
QY 301 GTNCCNCCNYTNCNCCNNGGNTTYCCNTTYGTNCCNCCNWSNNTYTYTYSNGCN 360
Db 427 GTTCCCTCCTCCTCCTAGGGGGTTTCCCGGTGTGTCCTTCACAGGTTTTTTTCAGCA 486

```

```

QY 361 GCNCGNCCNCCNCCNCCNCCNATHGCGNCGNARCCNCGNCGNCGNCCNNTNACN 420
Db 487 GCTGCAGCACCCGCTGCCCCACCTATTGACAGCTGAGCCTGCTGACGCTGCACCTTTACA 546
QY 421 GCNACNCCNCGNCGNCGNCGNCGNCGNCGNCGNCGNCGNCGNCGNCGNCGNCGN 480
Db 547 TCCACACCTGTAGCATCTGAGCCTGCTGCGAGGGGCCCTCTTGCAGCTGAGCCTGCTGCA 606
QY 481 GARGCNCCNCGNCGNCGNCGNCGNCGNCGNCGNCGNCGNCGNCGNCGNCGNCGN 540
Db 607 GAGCAGCTGTTGAGCTGAGCCTGCTGCGAGAGGACCTGTTGCAGCTGAGCCTGCTGCA 666
QY 541 GARGCNCCNCGNCGNCGNCGNCGNCGNCGNCGNCGNCGNCGNCGNCGNCGNCGN 600
Db 667 GAGCAGCTGTTGAGTGGAGCGCAGCTGCGAGAGGAACTTCACGAGCTGAGCCTGCTGCA 726
QY 601 GCNAARCCNCGNCGNCGNCGNCGNCGNCGNCGNCGNCGNCGNCGNCGNCGNCGN 657
Db 727 GCCAAGCCTGCTGCCCCAGAACCTCACCTTCTCCCTCTCTTGAACAGGCAAAATCAG 783

RESULT 5
AA61352
ID AAX61352 standard; cDNA; 959 BP.
XX
XX AAX61352;
XX
XX 14-JUL-1999 (first entry)
XX
XX DNA encoding a human secreted protein.
DE Human secreted protein; cancer; tumour; neurodegenerative disorder;
XX developmental abnormality; fetal deficiency; blood disorder; leukemia;
XX immune system disease; autoimmune disease; hepatic disease;
XX renal disease; lymphoma; inflammation; allergy; ischemic shock;
XX Alzheimer's; cognitive disorder; osteoporosis; schizophrenia; prostate disease;
XX obesity; osteoclast; osteoporosis; arthritis; malignancy; testes disease;
XX lung disease; thymus disease; digestive disorder; endocrine disorder;
XX infection; AIDS; ss.
XX Homo sapiens.
XX
XX WO9922243-Al.
XX
XX 06-MAY-1999.
XX
XX 23-OCT-1998; 98WO-US22376.
XX
XX 24-OCT-1997; 97US-0063387.
XX 24-OCT-1997; 97US-0062784.
XX 24-OCT-1997; 97US-0063088.
XX 24-OCT-1997; 97US-0063089.
XX 24-OCT-1997; 97US-0063090.
XX 24-OCT-1997; 97US-0063091.
XX 24-OCT-1997; 97US-0063092.
XX 24-OCT-1997; 97US-0063097.
XX 24-OCT-1997; 97US-0063098.
XX 24-OCT-1997; 97US-0063099.
XX 24-OCT-1997; 97US-0063100.
XX 24-OCT-1997; 97US-0063101.
XX 24-OCT-1997; 97US-0063109.
XX 24-OCT-1997; 97US-0063110.
XX 24-OCT-1997; 97US-0063111.
XX 24-OCT-1997; 97US-0063148.
XX 24-OCT-1997; 97US-0063386.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Brewer LA, Carter KC, Duan DR, Ebner R, Endress GA;
XX Feng P, Florence C, Florence JM, Janat F;
XX Kay H, Lafleur DW, Moore PA, Ni J, Olsen HS, Rosen CA;
XX Ruben SM, Shi Y, Soppet DR, Wei Y, Young P;
XX

```



```
SQ Sequence 296 BP; 50 A; 99 C; 82 G; 65 T; 0 other;
Query Match 30.1%; Score 197.8; DB 20; Length 296;
Best Local Similarity 62.7%; Pred. No. 1.2e-24;
Matches 185; Conservative 26; Mismatches 84; Indels 0; Gaps 0;

QY 322 GNTTYCCNTTGTGTCCTCCCTCAAGGTTTTTTCAGCAGCTGCAGCACCGCTGCCCA 381
DB 1 GGTTCGCTGTTGTCCTCCCTCAAGGTTTTTTCAGCAGCTGCAGCACCGCTGCCCA 60

QY 382 CCNATHGCGNGCARGCNGCNGCNGCNCNTTTCAGCAGCTGCAGCACCGCTGCCCA 441
DB 61 CCTATTGAGCTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 120

QY 442 CCNCGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNG 501
DB 121 CCGTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 180

QY 502 CCNCGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNG 561
DB 181 CCTGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 240

QY 562 CCNCGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNG 616
DB 241 CCAGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 295

RESULT 9
AAC84677
ID AAC84677 standard; DNA; 5120 BP.
AC AAC84677;
XX
XX
XX
DT 20-APR-2001 (first entry)
XX
XX
DE DNA sequence of human PPP2R2B.
KW Microsatellite marker; human; chromosome 5q31-33; D5S402; WI-6763;
KW CAG trinucleotide repeat; spinocerebellar ataxia 12; SCA12; PPP2R2Bbeta;
KW brain; protein phosphatase PP2A; ds.
XX
XX
OS Homo sapiens.
XX
XX
XX Key Location/Qualifiers
XX repeat_region 2088..2366
XX /rpt_type= "CAG"
XX /tag= a
XX mRNA 2473..3093
XX /tag= b
XX 5'UTR 2473..3023
XX /tag= C
XX CDS 3024..3093
XX /tag= d
XX /product= "PPP2R2B"
XX /note= "protein_id= AAF74024.1"
XX
XX
XX WO200078943-A2.
XX
XX 28-DEC-2000.
XX
XX 16-JUN-2000; 2000WO-US40213.
XX
XX 18-JUN-1999; 99US-0140176.
XX
XX (UYJO ) UNIV JOHNS HOPKINS.
XX
XX Margolis RL, Ross CA, Holmes SE;
XX
XX WPI; 2001-102718/11.
XX P-FSDB; AAB48340.
XX
XX New polynucleotide for use in diagnosing spinocerebellar ataxia 12,
XX comprises a microsatellite marker having a variable number of CAG
XX
```

```
PT trinucleotide repeats located on human chromosome 5q31-33 -
XX
XX Disclosure; Page 21-23; 23pp; English.
XX
XX The invention relates to a microsatellite marker located on human
XX chromosome 5q31-33 between markers D5S402 and WI-6763, comprising a
XX variable number of CAG trinucleotide repeats. The marker is useful for
XX diagnosing spinocerebellar ataxia 12 (SCA12) by determining the number of
XX CAG trinucleotide repeats in a 5'-untranslated region of an allele of
XX PPP2R2Bbeta, a brain specific regulatory subunit of protein phosphatase
XX PP2A, where if a number greater than 40 is determined, an expansion which
XX is associated with SCA12 is indicated. The presence of expanded
XX trinucleotide repeats on chromosome 5q31-33 is useful for predicting or
XX diagnosing SCA12. Also provided are primers useful for amplifying the
XX microsatellite marker associated with SCA12 and for diagnosing SCA12 and
XX a probe useful for detecting expansions in a microsatellite marker
XX located on human chromosome 5q31-33. The present sequence represents
XX a PPP2R2B DNA sequence.
XX
XX Sequence 5120 BP; 1336 A; 1267 C; 1268 G; 1249 T; 0 other;
XX
XX Query Match 15.8%; Score 103.8; DB 22; Length 5120;
XX Best Local Similarity 44.4%; Pred. No. 1e-08;
XX Matches 127; Conservative 21; Mismatches 138; Indels 0; Gaps 0;

QY 356 SNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNG 415
DB 2097 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2156

QY 416 TNACNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNG 475
DB 2157 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2216

QY 476 CNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNG 535
DB 2217 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2276

QY 536 CNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNG 595
DB 2277 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2336

QY 596 CNACNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNG 641
DB 2337 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2382

RESULT 10
AAA30290/C
ID AAA30290 standard; DNA; 3489 BP.
XX
XX AAA30290;
XX
XX 11-SEP-2000 (first entry)
XX
XX Kaposi's sarcoma-associated herpesvirus LANA gene.
XX
XX Kaposi's sarcoma-associated herpesvirus; KSHV; rhadino virus;
XX latency-associated nuclear antigen; LANA; gamma-2 herpes virus;
XX human herpes virus 8; HHV8; rhadino virus cis-acting element; RVCAE;
XX Kaposi's sarcoma; primary effusion lymphoma; PEL;
XX human immunodeficiency virus; HIV; multicentric Castlemann's disease; ds.
XX
XX Kaposi's sarcoma-associated herpesvirus.
XX
XX Key Location/Qualifiers
XX CDS 1..3489
XX /tag= a
XX /product= "LANA"
XX misc_signal 40..50
XX /tag= b
XX /note= "nuclear localisation signal, NLS"
XX misc_signal 190..210
XX /tag= c
XX
```


Db 1653 CTGCTGTGGCTCCTCTCTCTGTGGCTCTCTGTGGCTCCTCTCTCTGTGGCTCTCTG 1594

QY 596 CNACNGCNAARCCNGCNCNGARCC 623

Db 1593 CTGCTGTGGCTCCTCTCTCTGTGGCTCC 1566

RESULT 12

ABA93487/C

ID ABA93487 standard; DNA; 3489 BP.

XX ABA93487;

DT 25-APR-2002 (first entry)

XX Kaposi's sarcoma-associated herpesvirus LANA protein encoding DNA.

XX Kaposi's sarcoma-associated herpesvirus; KSHV; LANA; RVCAE; PEL;

KW KSHV terminal repeat; rhadino virus cis acting element; episome;

KW primary effusion lymphoma; latency-associated nuclear antigen;

KW gene therapy; gene transfer; gene; ds.

XX Human herpesvirus 8.

XX Key Location/Qualifiers

FT CDS 1..3489

FT /*tag= a

FT /product= "LANA protein"

FT /note= "latency-associated nuclear antigen"

XX US6322792-B1.

PN 27-NOV-2001.

XX 21-APR-1999; 99US-0298568.

XX 19-NOV-1998; 98US-109422P.

PR (KIEFF/) KIEFF E D.

XX Kieff ED, Ballestas ME, Kaye KW;

XX WPI; 2002-153769/20.

DR P-PSDB; ABB05621.

XX System for episomal retention of plasmids in mammalian cells, useful in gene therapy, comprises rhadinoviral LANA and RVCAE sequences -

PS Claim 1; Fig 6; 27pp; English.

XX The present invention describes a system (A) for maintaining a plasmid as an episome in mammalian cells, comprising the rhadinoviral sequence LANA (latency-associated nuclear antigen) of 3489 base pairs (see ABA93487, S1) expressed in the cell, and the rhadinoviral sequence RVCAE (rhadinoviral cis-acting element) of 801 base pairs (see ABA93488, S2) present in the plasmid. Also describes a method for maintaining a closed circular DNA in a cell by expressing (S1) in the cells and having (S2) as a cis-acting and maintenance sequence in the DNA. (A) is particularly used in gene therapy (or other gene transfer applications) that uses mammalian cells in which LANA is expressed. (A) improves persistence of gene therapy vectors in cells. The present sequence encodes Kaposi's sarcoma-associated herpesvirus (KSHV, also called human herpesvirus 8) LANA protein, which is used in the exemplification of the present invention.

XX Sequence 3489 BP; 1053 A; 862 C; 1137 G; 437 T; 0 other;

QY Query Match 15.2%; Score 99.6; DB 24; Length 3489;

Best Local Similarity 48.5%; Pred. No. 4.8e-08;

Matches 130; Conservative 7; Mismatches 131; Indels 0; Gaps 0;

QY 356 SNGCNGCNGCNCNCNCNCNCNATHGCGNGCNGARCCNGCNCNCNCNY 415

Db 1833 CTGCTGTCATCTATCTGCTGCTCATCTCTGCTCATCTGCTGCTGCTCCCG 1774

QY 416 TNACNGCNAACNCNGTNGCNGCNGARCCNGCNGCNGCNGCNGCNGARCCNG 475

Db 1773 CTGCTGTGGCTCCCGCTGCTGTGGCTCCCGCTGCTGTGGCTGCTGTGGCTCCCG 1714

QY 476 CNGCNGARCCNCNGTNGCNGCNGARCCNGCNGCNGCNGCNGCNGARCCNG 535

Db 1713 CTGCTGTGGCTCCCGCTGCTGTGGCTCCCGCTGCTGTGGCTGCTGTGGCTCCCG 1654

QY 536 CNGCNGARCCNCNGTNGCNGCNGARCCNGCNGCNGCNGCNGCNGCNGARCCNG 595

Db 1653 CTGCTGTGGCTCCTCTCTCTGTGGCTCCTCTCTGTGGCTCCTCTGTGGCTCCTCTG 1594

QY 596 CNACNGCNAARCCNGCNCNGARCC 623

Db 1593 CTGCTGTGGCTCCTCTCTCTGTGGCTCC 1566

RESULT 13

AAV73805

ID AAV73805 standard; DNA; 32207 BP.

XX AAV73805;

XX 25-FEB-1999 (first entry)

XX KSHV LUR DNA (nucleotides 105,301-137,507).

DE Kaposi's sarcoma; acquired immune deficiency syndrome; AIDS; DHFR; Bcl-2;

XX dihydrofolate reductase; LUR; long unique region; vaccine; prophylaxis;

KW diagnosis; treatment; HHV8; capsid protein IV; tegument protein IV;

KW glycoprotein; kaposin; cyclin D; immediate early protein; IEP; OX-2;

KW v-ah; G-protein coupled receptor; FGARAT; ds.

XX Kaposi's sarcoma-associated herpesvirus.

OS US5849564-A.

XX 15-DEC-1998.

XX 29-NOV-1996; 96US-0770379.

XX 29-NOV-1996; 96US-0770379.

XX (UYCO) UNIV COLUMBIA NEW YORK.

PA Bohenzky RA, Chang Y, Edelman IS, Moore PS, Russo JJ;

XX WPI; 1999-069741/06.

DR Kaposi's sarcoma-associated herpes virus nucleic acid - encodes

XX di:hydro:folate reductase and is useful for treatment, prophylaxis

PT or diagnosis of Kaposi's sarcoma

XX Disclosure; Column 155-182; 109pp; English.

XX This sequence is a fragment of the Kaposi's sarcoma-associated

CC herpesvirus (KSHV) LUR (long unique region). This fragment contains

CC coding regions for ORF65 which encodes capsid protein IV, ORF66, ORF67

CC which encodes tegument protein IV, ORF68 which encodes a glycoprotein,

CC ORF69, K12 which encodes kaposin, K13, ORF72 which encodes cyclin D,

CC ORF73 which encodes immediate early protein (IEP), K14 which encodes

CC OX-2 (v-ah), ORF74 which encodes G-protein coupled receptor, ORF75

CC which encodes tegument protein/FGARAT, K15. KSHV is a new human

CC Herpesvirus (HHV8) believed to cause Kaposi's sarcoma (KS) which is the

CC most common form of neoplasm occurring in persons with acquired immune

CC deficiency syndrome (AIDS). The DHFR protein is useful for vaccination,

CC prophylaxis, diagnosis and treatment of a subject with Kaposi's sarcoma

CC and for detecting expression of a DNA virus associated with Kaposi's

XX sarcoma in a cell.

SQ Sequence 32207 BP; 7229 A; 9156 C; 8713 G; 7109 T; 0 other;

Query Match 15.2%; Score 99.6; DB 20; Length 32207;
Best Local Similarity 48.5%; Pred. No. 6.7e-08;
Matches 130; Conservative 7; Mismatches 131; Indels 0; Gaps 0;

QY 356 SNGCNGCNGCNGCNGCNGCNGCCNCCNATHGCGNCNGARCCNGCNGCNGCNGCNGCNY 415
: ||| : ||| ||| ||| : ||| ||| ||| ||| ||| ||| |||

Db 20164 CTGCTGCTCATCTGCTGCTGCATCTGCTGCTGCATCTGCTGCTGCTGCCG 20223

QY 416 TNACNCGNACNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNG 475
: ||| : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Db 20224 CTGCTGTGGCTCCCGCTGTGTGGCTCCCGCTGTGTGGCTCCCGCTGTGTGGCTCCCG 20283

QY 476 CNGCNGARGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNG 535
||| : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Db 20284 CTGCTGTGGCTCCCGCTGTGTGGCTCCCGCTGTGTGGCTCCCGCTGTGTGGCTCCCG 20343

QY 536 CNGCNGARGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNG 595
||| : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Db 20344 CTGCTGTGGCTCCCGCTGTGTGGCTCCCGCTGTGTGGCTCCCGCTGTGTGGCTCCCG 20403

QY 596 CNACNCNGAACCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNG 623
||| : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Db 20404 CTGCTGTGGCTCCCGCTGTGTGGCTCCCGCTGTGTGGCTCCCGCTGTGTGGCTCCCG 20431

RESULT 14

AAV19941

ID AAV19941 standard; DNA; 137507 BP.

XX AAV19941;

XX 03-AUG-1998 (first entry)

XX KSHV long unique coding region and terminal repeat.

XX KSHV; HHV8; human herpes virus 8; macrophage inflammatory protein II; interleukin-6; IL-6; interferon regulatory factor; rheumatoid arthritis; complement-binding protein; glycoprotein; capsid protein IV; infection; immediate early protein; Kaposi's sarcoma; protective vaccine; lymphoma; lymphoproliferative disease; leukaemia; splenomegaly; mycosis fungoides; HIV immune status; anti-inflammatory agent; therapy; ds.

XX Kaposi's sarcoma-associated herpes virus.

XX Key Location/Qualifiers

CDS 1142..2794

FT /*tag= a

FT /product= complement-binding protein

CDS 8699..11236

FT /*tag= b

FT /product= glycoprotein B

CDS complement (17261..17875)

FT /*tag= c

FT /product= interleukin 6

CDS complement (21548..21832)

FT /*tag= d

FT /product= macrophage inflammatory protein II

CDS complement (27137..27424)

FT /*tag= e

FT /product= interferon regulatory factor 1

CDS 28661..29741

FT /*tag= f

FT /product= protein Tl.1

CDS complement (58976..60175)

FT /*tag= g

FT /product= glycoprotein M

CDS complement (69412..69915)

FT /*tag= h

FT /product= glycoprotein L

CDS complement (88410..88910)

FT /*tag= i

/product= interferon regulatory factor 2

89600..90541

/*tag= j

/product= interferon regulatory factor 3

90173..90643

/*tag= k

/product= glycoprotein X

complement (93636..94127)

/*tag= l

/product= interferon regulatory factor 4

complement (111931..112443)

/*tag= m

/product= capsid protein IV

complement (123808..127296)

/*tag= n

/product= immediate early protein

WO9804576-A1.

05-FEB-1998.

22-JUL-1997; 97WO-US13346.

29-NOV-1996; 96US-0757669.

25-JUL-1996; 96US-0686243.

25-JUL-1996; 96US-0686349.

25-JUL-1996; 96US-0686350.

25-JUL-1996; 96US-0687253.

25-JUL-1996; 96US-0688814.

05-SEP-1996; 96US-0708678.

10-OCT-1996; 96US-0728323.

13-NOV-1996; 96US-0747887.

13-NOV-1996; 96US-0748640.

(UVC) UNIV COLUMBIA NEW YORK.

Bohenzky RA, Chang Y, Edelman IS, Moore PS, Russo JJ;

WPI; 1998-130615/12.

New nucleic acid encoding Kaposi's sarcoma associated herpes virus proteins - useful for, e.g. detecting levels of HHV8 in, and preparation of vaccines for treatment of, HIV patients

Example 2; Page 135-203; 230pp; English.

This sequence represents the long unique region and terminal repeat of the Kaposi's sarcoma-associated herpes virus (KSHV). KSHV is also known as human herpes virus 8 (HHV8). This sequence contains the DNAs of the invention which encode KSHV polypeptides selected from: (a) viral macrophage inflammatory protein (MIP) II; (b) viral interleukin-6 (IL-6); (c) viral IRF 1; (d) complement-binding protein; glycoproteins B, M or L; (e) capsid protein IV encoded by ORF65; and (f) immediate early protein encoded by ORF73. Labelled probes for the nucleic acid, proteins encoded by it, and antibodies (Ab) specific for the proteins are useful for detecting HHV8, specifically for diagnosis of Kaposi's sarcoma, in body fluids or tissue samples. HHV8 infections can be treated with antiseptic or triplex forming molecules or agents that bind specifically to the protein. Ab may be used for prophylaxis or treatment of HHV8 infection, while the protein can be used in protective vaccines. Ab may also be used to differentiate between lymphomas, and HHV8 may be implicated in many other lymphoproliferative diseases such as lymphomas, leukaemia, splenomegaly and mycosis fungoides. Cells and animals containing the nucleic acid are useful for drug screening. HHV8-derived peptides can be used as targets for antiviral drugs, e.g. dihydrofolate reductase gene can be inhibited with methotrexate. These can also be used to determine the immune status of a patient infected with HIV. HHV8 derived protein viral MIP III may be used as an anti-inflammatory agent for, e.g. treating rheumatoid arthritis. This sequence is stated as containing 81 open reading frames.

Sequence 137507 BP; 32579 A; 37795 C; 35758 G; 31375 T; 0 other;


```
; APPLICANT: Orr, Harry T.
; APPLICANT: Ranum, Laura P.W.
; APPLICANT: Chung, Ming-yi
; APPLICANT: Zoghbi, Huda Y.
; TITLE OF INVENTION: Gene Sequence for Spinocerebellar Ataxia
; Patent No. 5834183
; TITLE OF INVENTION: Type 1 and Method for Diagnosis
; NUMBER OF SEQUENCES: 85
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Muehling, Raasch, Gebhardt & Schwappach, P.A.
; STREET: P.O. Box 581415
; CITY: Minneapolis
; STATE: MN
; COUNTRY: USA
; ZIP: 55458-1415
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/267,803B
; FILING DATE: 28-JUN-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: McCormack, Myra H.
; REGISTRATION NUMBER: 36,602
; REFERENCE/DOCKET NUMBER: 110.00030120
; TELEPHONE: 612-305-1217
; TELEFAX: 612-305-1228
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; TYPE: nucleic acid
; LENGTH: 234 base pairs
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
; US-08-267-803B-3

Query Match 13.1%; Score 86.2; DB 2; Length 234;
Best Local Similarity 45.7%; Pred. No. 5.8e-10;
Matches 107; Conservative 14; Mismatches 113; Indels 0; Gaps 0;

QY 359 CNGCGNCNCNGCNGCNGCNCNCNATHGCGNGCNGARCCNGCNGCNCNCNYTNA 418
Db 1 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 60
QY 419 CNGCNCNCNCNGCNGCNGARCCNGCNGCNGCNGCNGCNGCNGCNGCNGCNG 478
Db 61 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 120
QY 479 CNGARCCNCNCNGCNGCNGARCCNGCNGCNGCNGCNGCNGCNGCNGCNGCNG 538
Db 121 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 180
QY 539 CNGARCCNCNCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNG 592
Db 181 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 234

RESULT 12
5482709-5
; Patent No. 5482709
; APPLICANT: JACOBSON, JAMES W.; STRAUSBERG, ROBERT L.; WILSON,
; SUSAN D.; POPE, SHARON H.; STRAUSBERG, SUSAN L.; RUFF, MICHAEL D.;
; AUGUSTINE, PATRICIA C.; DANFORTH, HARRY D.
; TITLE OF INVENTION: EIMERIA ANTIGENIC COMPOSITION WHICH
; ELICITS ANTIBODIES AGAINST AVIAN COCCIDIOSIS
; NUMBER OF SEQUENCES: 10
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/148,432
; FILING DATE: 08-NOV-1993

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 581,693
; FILING DATE: 12-SEP-1990
; APPLICATION NUMBER: 215,162
; FILING DATE: 05-JUL-1989
; APPLICATION NUMBER: 746,520
; FILING DATE: 19-JUN-1985
; APPLICATION NUMBER: 627,811
; FILING DATE: 05-JUL-1984
; SEQ ID NO: 5:
; LENGTH: 533
; 5482709-5

Query Match 13.0%; Score 85.2; DB 6; Length 533;
Best Local Similarity 39.2%; Pred. No. 1.3e-09;
Matches 115; Conservative 25; Mismatches 153; Indels 0; Gaps 0;

QY 343 WSNMGNTTYTYWSNGCNGCNGCNCNCNCNCNCNCNATHGCGNGCNGARCCNGC 402
Db 168 AGCAGCAGCTACACTGCAGCTGCAGCAGCAGCAGCTGCAGCAGCAGCAGCTGCAGCAGCA 227
QY 403 GCNGCNGCNCNYTNACNGCNCNCNCNGTNGCNGCNGARCCNGCNGCNGCNGCNGTGN 462
Db 228 GCAGCAACAGGAGCTGCAGCAGCAGCAGCTGCAGCAGCAGCAGCAGCAGCAGGGA 287
QY 463 GCNGCNGARCCNGCNGCNGARCCNCNCNGTNGCNGCNGARCCNGCNGCNGCNGCNGTGN 522
Db 288 GCACCAAGAGCAGCGGAAGCAGCAGCAACAGAAACCGAAGGAGCAGCAGCAGCAGAA 347
QY 523 GCNGCNGARCCNGCNGCNGARCCNCNCNGTNGCNGCNGARCCNGCNGCNGCNGARCCNWSN 582
Db 348 GCAGAGAAAGCGAAACACAGCAGCAGCAACAGCAGCAACACAGCAGCTGCAGCAAGAGCA 407
QY 583 CCNGCNGARCCNGCNCNACNGCNCNCNCNGCNGCNGCNGCNGCNGCNGCNGCNGCNG 635
Db 408 GCAGCAGCATCAGCAGCAGGTAAAGCCTGGGGGACATGCAGCAGCATCAACAGC 460

RESULT 13
US-07-814-220-3
; Sequence 3, Application US/07814220
; Patent No. 5925540
; GENERAL INFORMATION:
; APPLICANT: Caceci, Thomas
; APPLICANT: Toth, Thomas E.
; APPLICANT: Szumanski, Maria B.W.
; TITLE OF INVENTION: SYNTHETIC ANTIFREEZE PEPTIDE AND
; TITLE OF INVENTION: SYNTHETIC GENE CODING FOR ITS PRODUCTION
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: WHITHAM, CURTIS & WHITHAM
; STREET: Reston Intl. Center, 11800 Sunrise Valley Dr.,
; CITY: Reston
; STATE: VA
; COUNTRY: USA
; ZIP: 20191
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/814,220
; FILING DATE: 23-DEC-1991
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/588,437
; FILING DATE: 25-SEP-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Whitham, Michael E.
; REGISTRATION NUMBER: 32,635
; REFERENCE/DOCKET NUMBER: CIT.016
```



```
; REFERENCE/DOCKET NUMBER: CIT.016
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-391-2510
; TELEFAX: 703-391-9035
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 336 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 31..324
; US-07-812-421-3

Query Match      12.7%; Score 83.4; DB 2; Length 336;
Best Local Similarity 41.4%; Pred. No. 2.6e-09;
Matches 118; Conservative 16; Mismatches 151; Indels 0; Gaps 0;

QY 341 CNWSNMGNFTYTYWSNGCNGCNGCNGCNGCNCNCNATHGCGCNGGARGCCNG 400
   | : : | : : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 35 CCACCCGCTCCGACGCGCGCGGAGCAGCTGCTACTGCGGCGACCCGACGAGCGGCTG 94
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 401 CNGCNGCNGCNCNNTYTNACNGNACNCCNGTNGCNGCNGARGCCNGCNGGNGCNCNG 460
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 95 CGGCAGCTACTGCGGCGACCGCAGCAGCGGCTGCGGCAGCTACTGCGGCGACCGCAGCAG 154
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 461 TNGCNGCNGARGCCNGCNGCNGARGCNCNCCNGTNGGNGCNGARGCCNGCNGARGCNCNG 520
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 155 CGGCTGCGGCGAGCTACCGGCTGTACCGGAGCAGCTGCTGCGGCGAGCAACTGCGGCGACCG 214
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 521 TNGCNGCNGARGCCNGCNGCNGARGCNCNCCNGTNGGNGTNGARGCCNGCNGARGARCCNW 580
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 215 CAGCAGCGGCTCGGCGAGCTACCGCAGCTACTGCAGCTAAACCTCGGCTCTGACTGCTG 274
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 581 SNGCNGCNGARGCCNGCNGCNGNACNAAACCCNGCNGCNGCNGCNGCNGCNC 625
   : | | : | | : | | : | | : | | : | | : | | : | | : | |
Db 275 CAAACGACGCGGCTGTGCGGCGGAGCTGTGCTGCGGCGAGCGGCTC 319
```

Search completed: February 12, 2003, 11:57:18
Job time : 88.2486 secs


```

Db 254 CCTGCTCCGCCCGCCAGCTGCTGTGATTTCCAGTTGCTGCGAGGCTTCTCTGTATCTCCA 313
QY 460 GTNGCNGCNGARCCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGC 519
Db 314 GCTGCTCAAAACCCAGTGTGTCAGACACACCTGCTGCGCCCGCAGCTGCTGCAATTTCCA 373
QY 520 GTNGCNGCNGARCCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGC 579
Db 374 GTTGCTGCAGGCTTCTGCTGTATCTCCAGCTGCTGCAAAACCCAGCTGCTGCAGGACCA 433
QY 580 WSNCCNGCNGARCCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGC 629
Db 434 CCTGCTCCGCCCGCCAGCTGCTGTGATTTAGTTGCTGCGAGGCTTCTCTGC 483

RESULT 8
US-09-880-107-3699
; Sequence 3699, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3699
; LENGTH: 6483
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 X14487
US-09-880-107-3699

Query Match 12.5%; Score 81.8; DB 10; Length 6483;
Best Local Similarity 36.6%; Pred. No. 1.2e-08;
Matches 141; Conservative 23; Mismatches 221; Indels 0; Gaps 0;

QY 251 TYCCNTAYGNTAYCAYATHMNGGNTTYCCNTTNCNACNCARYTNAAYGTNCCNCNY 310
Db 5044 TACCCACCGTTTTCACACGTTAGAGATAACCGAATGTGACCTCACCCCGTTTGTTCGG 5103
QY 311 TNCNCCNMGNGNTTYCCNTTYTNCNCCNWSNMGNTTYTWSNGCNGCNGCNGCNC 370
Db 5104 AGCGGGGACGGCGGGCGGAAGTTTCGGCGCGGCTACGGCGGGCGGAAGCTCCGGCGG 5163
QY 371 CNGCNGCNCNCNATHGNGCNGARCCNGCNGCNGCNGCNCNTYNACNGCNCACNG 430
Db 5164 CGGAAGCTCGGGCGGGGCTACGGCGGGCGGCACGGCGGAGTTCCGGCGGGCTACGG 5223
QY 431 TNGCNGCNGARCCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGC 490
Db 5224 AGCGGAAAGCTCCGGCGGGCGGAAGCTCCGGCGGGCGGCTACGGGGGCGGAAGCTCCAGCG 5283
QY 491 TNGGNGCNGARCCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGC 550
Db 5284 CGGCCACGGGGCGGAAGTTCAGCGGGCGGCACGGCGGAGTTCCAGCGGGGCTACGG 5343
QY 551 TNGGNGTNGARCCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGC 610
Db 5344 TGGTGCAAGTTCCGGCGGGCGGGCGGGCTACGGGGGCGGAGCTCCGGCGGGCGGAG 5403
QY 611 CNGCNGCNGARCCNCAYCCNWSNCC 635
; APPLICANT: Steininger II, Robert J.
```

```

Db 5404 CAGCTCCGGCGCGGATACGCGGC 5428

RESULT 9
US-09-880-107-1748
; Sequence 1748, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1748
; LENGTH: 6604
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 D83783
US-09-880-107-1748

Query Match 12.4%; Score 81.2; DB 10; Length 6604;
Best Local Similarity 38.2%; Pred. No. 1.7e-08;
Matches 120; Conservative 23; Mismatches 171; Indels 0; Gaps 0;

QY 337 CCNCCNWSNMGNTTYTWSNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGAR 396
Db 5967 CCGTTCAACAGCCATCTCTACCTGAGCAGCAGCAGCAACAGCAGCAACA 6026
QY 397 CCNGCNGCNGCNGCNCNTTACNGCNCNACNGCNGCNGCNGCNGCNGCNGCNGCNG 456
Db 6027 GCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 6086
QY 457 CCNGTNGCNGCNGARCCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNG 516
Db 6087 GCAGCAGCAGCAGATCTCTCGCGCAGCAGCAGCAGCAACAGCAGCAGCAGCAGCAGCAG 6146
QY 517 CCNGTNGCNGCNGARCCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGTNGCNGCNGCNGAR 576
Db 6147 GCAACAGCAACAGCAGCAGCAGCAGCAACAACAACAGCAGCAGCAGCAACAGCAGCA 6206
QY 577 CCNWSNCCNGCNGARCCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNG 636
Db 6207 ACAGCGGCTCTCTCCCAACCCCGCCCTGCCAGCCCGGTTCCAGCGCCAGGGGCT 6266
QY 637 WSNYTNAGARCC 650
Db 6267 TCAGCAGACCCAGC 6280

RESULT 10
US-09-822-846-491
; Sequence 491, Application US/09822846
; Publication No. US20030027139A1
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John M.
; APPLICANT: LaVallie, Edward R.
; APPLICANT: Collins-Racie, Lisa A.
; APPLICANT: Evans, Cheryl
; APPLICANT: Merberg, David
; APPLICANT: Treacy, Maurice
; APPLICANT: Agostino, Michael J.
; APPLICANT: Steininger II, Robert J.
```


Db 1554 GCAGCAGCTCCGGCGCGGATACGGCGGCGAGCAGCTCCGGCGGCGGATACGGCGCG 1613
QY 598 ACNGCNAARCCNGCNGCNGARCCNCAVC 628
Db 1614 GCAGCTCAGCGGAGGCCACACAGTCTCTCTC 1644

RESULT 13
US-09-976-740-48
; Sequence 48, Application US/09976740
; Publication No. US20020194633A1
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
; TITLE OF INVENTION: ATHEROSCLEROSIS
; FILE REFERENCE: 10797-004001
; CURRENT APPLICATION NUMBER: US/09/976,740
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 09/616,289
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 08/979,608
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 60/031,930
; PRIOR FILING DATE: 1996-11-27
; PRIOR APPLICATION NUMBER: US 60/048,547
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 48
; LENGTH: 2561
; TYPE: DNA
; ORGANISM: Oryctolagus cuniculus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (246)...(1895)
US-09-976-740-48

Query Match 11.6%; Score 76.4; DB 9; Length 2561;
Best Local Similarity 35.5%; Pred. No. 1.2e-07;
Matches 132; Conservative 25; Mismatches 215; Indels 0; Gaps 0;
QY 286 GCNACNARYTNAAYGTNCCNCCNTNCCNMGNGNTTYCNCNTTYGTNCCNCCNWSN 345
Db 564 GCCACCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 623
QY 346 MGNNTTYTWSNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNG 405
Db 624 CGGCCACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 683
QY 406 GCNGCNCNNTNACNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNG 465
Db 684 CGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 743
QY 466 GCNGARCCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNG 525
Db 744 GCGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 803
QY 526 GCNGARCCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNG 585
Db 804 GCGCGTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 863
QY 586 GCNGARCCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNG 645
Db 864 GCGCGTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 923
QY 646 CARGCNAAYCAR 657
Db 924 CAGCAGCAGCAG 935

RESULT 15
US-10-023-523-48
; Sequence 48, Application US/10023523

RESULT 14
US-10-023-529-48
; Sequence 48, Application US/10023529
; Patent No. US20020129388A1
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
; TITLE OF INVENTION: ATHEROSCLEROSIS
; FILE REFERENCE: 10797-004001
; CURRENT APPLICATION NUMBER: US/10/023,529
; CURRENT FILING DATE: 2001-12-17
; PRIOR APPLICATION NUMBER: 09/616,289
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 09/517,849
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 08/979,608
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 60/031,930
; PRIOR FILING DATE: 1996-11-27
; PRIOR APPLICATION NUMBER: US 60/048,547
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 48
; LENGTH: 2561
; TYPE: DNA
; ORGANISM: Oryctolagus cuniculus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (246)...(1895)
US-10-023-529-48

Query Match 11.6%; Score 76.4; DB 12; Length 2561;
Best Local Similarity 35.5%; Pred. No. 1.2e-07;
Matches 132; Conservative 25; Mismatches 215; Indels 0; Gaps 0;
QY 286 GCNACNARYTNAAYGTNCCNCCNTNCCNMGNGNTTYCNCNTTYGTNCCNCCNWSN 345
Db 564 GCCACCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 623
QY 346 MGNNTTYTWSNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNG 405
Db 624 CGGCCACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 683
QY 406 GCNGCNCNNTNACNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNG 465
Db 684 CGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 743
QY 466 GCNGARCCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNG 525
Db 744 GCGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 803
QY 526 GCNGARCCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNG 585
Db 804 GCGCGTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 863
QY 586 GCNGARCCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNG 645
Db 864 GCGCGTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 923
QY 646 CARGCNAAYCAR 657
Db 924 CAGCAGCAGCAG 935

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 12, 2003, 11:44:27 ; Search time 1604.42 Seconds
(without alignments)
10295.682 Million cell updates/sec

Title: US-09-923-236-3
Perfect score: 657
Sequence: 1 atgaarytnytnytnynggc.....snyngarcargcnaaycar 657

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 24791104 seqs, 12571243825 residues

Total number of hits satisfying chosen parameters: 49582208

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_NA_Main:*

1: /cgn2_6/ptodata/1/pna/PCTUS_COMB.seq:*

2: /cgn2_6/ptodata/1/pna/US06_COMB.seq:*

3: /cgn2_6/ptodata/1/pna/US07_COMB.seq:*

4: /cgn2_6/ptodata/1/pna/US080_COMB.seq:*

5: /cgn2_6/ptodata/1/pna/US081_COMB.seq:*

6: /cgn2_6/ptodata/1/pna/US082_COMB.seq:*

7: /cgn2_6/ptodata/1/pna/US083_COMB.seq:*

8: /cgn2_6/ptodata/1/pna/US084_COMB.seq:*

9: /cgn2_6/ptodata/1/pna/US085_COMB.seq:*

10: /cgn2_6/ptodata/1/pna/US086_COMB.seq:*

11: /cgn2_6/ptodata/1/pna/US087_COMB.seq:*

12: /cgn2_6/ptodata/1/pna/US088_COMB.seq:*

13: /cgn2_6/ptodata/1/pna/US089_COMB.seq:*

14: /cgn2_6/ptodata/1/pna/US090_COMB.seq:*

15: /cgn2_6/ptodata/1/pna/US091_COMB.seq:*

16: /cgn2_6/ptodata/1/pna/US092_COMB.seq:*

17: /cgn2_6/ptodata/1/pna/US093_COMB.seq:*

18: /cgn2_6/ptodata/1/pna/US094_COMB.seq:*

19: /cgn2_6/ptodata/1/pna/US095A_COMB.seq:*

20: /cgn2_6/ptodata/1/pna/US095B_COMB.seq:*

21: /cgn2_6/ptodata/1/pna/US095C_COMB.seq:*

22: /cgn2_6/ptodata/1/pna/US095D_COMB.seq:*

23: /cgn2_6/ptodata/1/pna/US096A_COMB.seq:*

24: /cgn2_6/ptodata/1/pna/US096B_COMB.seq:*

25: /cgn2_6/ptodata/1/pna/US096C_COMB.seq:*

26: /cgn2_6/ptodata/1/pna/US096D_COMB.seq:*

27: /cgn2_6/ptodata/1/pna/US096E_COMB.seq:*

28: /cgn2_6/ptodata/1/pna/US097A_COMB.seq:*

29: /cgn2_6/ptodata/1/pna/US097B_COMB.seq:*

30: /cgn2_6/ptodata/1/pna/US097C_COMB.seq:*

31: /cgn2_6/ptodata/1/pna/US098A_COMB.seq:*

32: /cgn2_6/ptodata/1/pna/US098B_COMB.seq:*

33: /cgn2_6/ptodata/1/pna/US098C_COMB.seq:*

34: /cgn2_6/ptodata/1/pna/US099A_COMB.seq:*

35: /cgn2_6/ptodata/1/pna/US099B_COMB.seq:*

36: /cgn2_6/ptodata/1/pna/US099C_COMB.seq:*

37: /cgn2_6/ptodata/1/pna/US099D_COMB.seq:*

38: /cgn2_6/ptodata/1/pna/US100A_COMB.seq:*

39: /cgn2_6/ptodata/1/pna/US100B_COMB.seq:*

40: /cgn2_6/ptodata/1/pna/US101A_COMB.seq:*

41: /cgn2_6/ptodata/1/pna/US101B_COMB.seq:*

42: /cgn2_6/ptodata/1/pna/US102A_COMB.seq:*

43: /cgn2_6/ptodata/1/pna/US102B_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	460	70.0	657	34	US-09-922-469-3
2	460	70.0	657	34	US-09-922-480-3
3	460	70.0	657	34	US-09-923-236-3
4	460	70.0	961	1	PCT-US02-08123-144
5	460	70.0	961	1	PCT-US02-08277-108
6	460	70.0	961	1	PCT-US02-08278-143
7	460	70.0	998	38	US-10-009-328-72
8	460	70.0	998	60	US-60-162-287-23
9	460	70.0	1002	76	US-60-324-185-219
10	460	70.0	1008	34	US-09-922-469-1
11	460	70.0	1008	34	US-09-922-480-1
12	460	70.0	1008	34	US-09-923-236-1
13	456.8	69.5	1325	1	PCT-US98-13813-9
14	456.8	69.5	1325	13	US-08-958-304-9
15	456.8	69.5	1325	22	US-09-746-783-105
16	449	68.3	956	22	US-09-585-715-10
17	449	68.3	956	57	US-60-137-261-9
18	449	68.3	979	22	US-09-585-715-149
19	449	68.3	979	60	US-60-167-410-37
20	429	65.3	959	1	PCT-US02-08123-695
21	429	65.3	959	1	PCT-US02-08277-471

22 429 65.3 959 1 PCT-US02-08278-666 Sequence 666, App
23 429 65.3 959 1 PCT-US98-22376-41 Sequence 41, Appl
24 429 65.3 959 16 US-09-296-622-41 Sequence 41, Appl
25 429 65.3 959 42 US-10-231-417-41 Sequence 41, Appl
26 419.8 63.9 783 25 US-09-649-163-10489 Sequence 10489, A
27 419.8 63.9 783 25 US-09-652-122-4871 Sequence 4871, Ap
28 419.8 63.9 783 29 US-09-726-170-2168 Sequence 2168, Ap
29 402.8 61.3 870 39 US-10-098-754-4114 Sequence 4114, Ap
30 384.6 58.5 695 25 US-09-652-122-2598 Sequence 2598, Ap
31 354.4 53.9 774 29 US-09-726-170-1856 Sequence 1856, Ap
32 322.6 49.1 621 25 US-09-652-122-2789 Sequence 2789, Ap
33 319 48.6 584 25 US-09-652-122-2510 Sequence 2510, Ap
34 314.2 47.8 574 18 US-09-496-911-12587 Sequence 12587, A
35 314.2 47.8 574 36 US-09-975-673A-12587 Sequence 4763, A
36 299.6 45.6 595 32 US-09-834-366-4763 Sequence 4763, Ap
37 299.6 45.6 595 63 US-60-197-873-4763 Sequence 1366, Ap
38 299.2 45.5 639 24 US-09-621-976-1366 Sequence 1366, Ap
39 299.2 45.5 639 58 US-60-147-499-1366 Sequence 12611, A
40 291.4 44.4 622 18 US-09-496-911-12611 Sequence 12611, A
41 291.4 44.4 622 36 US-09-975-673A-12611 Sequence 218, App
42 288 43.8 550 51 US-60-070-771-218 Sequence 218, Appl
43 268.8 40.9 420 29 US-09-726-170-82 Sequence 10956, A
44 267.8 40.8 396 16 US-09-233-972-10956 Sequence 10956, A
45 267.8 40.8 396 34 US-09-904-939-10956 Sequence 10956, A

ALIGNMENTS

RESULT 1
US-09-922-469-3
; Sequence 3, Application US/09922469
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; APPLICANT: Adler, David A.
; TITLE OF INVENTION: SECRETED SALIVARY ZSIG63 POLYPEPTIDE
; FILE REFERENCE: 97-71
; CURRENT APPLICATION NUMBER: US/09/922,469
; CURRENT FILING DATE: 2001-08-03
; PRIOR APPLICATION NUMBER: US 60/124,820
; PRIOR FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 657
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Degenerate polynucleotide sequence for zsig63
; NAME/KEY: misc_feature
; LOCATION: (1)...(657)
; OTHER INFORMATION: n = A,T,C or G
US-09-922-469-3

Query Match 70.0%; Score 460; DB 34; Length 657;
Best Local Similarity 100.0%; Pred. No. 1.5e-58;
Matches 657; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGAARYTNTYNTGGCGNTGYATHGTNTGYTGNGCNTTYGCMNGNAARMGNMNTTY 60
DB 1 ATGAARYTNTYNTGGCGNTGYATHGTNTGYTGNGCNTTYGCMNGNAARMGNMNTTY 60
QY 61 CCNTTYATHGGNGARGAYGAYAAAYGAYGGNCAVCNNTNCAVCNWSNTNAAAYATH 120
DB 61 CCNTTYATHGGNGARGAYGAYAAAYGAYGGNCAVCNNTNCAVCNWSNTNAAAYATH 120
QY 121 CCNTAYGGNATHMGNAAYTNCNCCNCCNNTNTATAYTMGNCCNGTNAAYACNGTNCN 180
DB 121 CCNTAYGGNATHMGNAAYTNCNCCNCCNNTNTATAYTMGNCCNGTNAAYACNGTNCN 180
QY 181 WSNATCCNGGNAAYACNTAYACNGAYACNGGNYTNCNWSNTAYCCNTGGATHYTACN 240
DB 181 WSNATCCNGGNAAYACNTAYACNGAYACNGGNYTNCNWSNTAYCCNTGGATHYTACN 240

QY 241 WSNCCNGGNTTYCCNTAYGTNTAYCAVATHMGNGNTTYCCNTNGCNACNARYTNAAY 300
DB 241 WSNCCNGGNTTYCCNTAYGTNTAYCAVATHMGNGNTTYCCNTNGCNACNARYTNAAY 300
QY 301 GTNCCNCCNTNCCNCCNMGGNTTYCCNTTYGTNCCNCCNWSNMGNMTTYTWSNGCN 360
DB 301 GTNCCNCCNTNCCNCCNMGGNTTYCCNTTYGTNCCNCCNWSNMGNMTTYTWSNGCN 360
QY 361 GCNCGNCCNCCNCCNCCNCCNATHGCGNGCGNGCGNGCGNGCGNGCGNGCGNACN 420
DB 361 GCNCGNCCNCCNCCNCCNCCNATHGCGNGCGNGCGNGCGNGCGNGCGNGCGNACN 420
QY 421 GCNACNCCNCGTNGCGNGCGNGCGNGCGNGCGNGCGNGCGNGCGNGCGNGCN 480
DB 421 GCNACNCCNCGTNGCGNGCGNGCGNGCGNGCGNGCGNGCGNGCGNGCGNGCN 480
QY 481 GARGCNCNCGTNGCGNGCGNGCGNGCGNGCGNGCGNGCGNGCGNGCGNGCGNGCN 540
DB 481 GARGCNCNCGTNGCGNGCGNGCGNGCGNGCGNGCGNGCGNGCGNGCGNGCGNGCN 540
QY 541 GARGCNCNCGTNGCGNGCGNGCGNGCGNGCGNGCGNGCGNGCGNGCGNGCGNGCN 600
DB 541 GARGCNCNCGTNGCGNGCGNGCGNGCGNGCGNGCGNGCGNGCGNGCGNGCGNGCN 600
QY 601 GCNAARCCNCGNCCNCCNCCNCAVCNWSNCCNWSNTNGARGCGNAAAYCAR 657
DB 601 GCNAARCCNCGNCCNCCNCCNCAVCNWSNCCNWSNTNGARGCGNAAAYCAR 657

RESULT 2
US-09-922-480-3
; Sequence 3, Application US/09922480
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; APPLICANT: Adler, David A.
; TITLE OF INVENTION: SECRETED SALIVARY ZSIG63 POLYPEPTIDE
; FILE REFERENCE: 97-71
; CURRENT APPLICATION NUMBER: US/09/922,480
; CURRENT FILING DATE: 2001-08-03
; PRIOR APPLICATION NUMBER: US 60/124,820
; PRIOR FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 657
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Degenerate polynucleotide sequence for zsig63
; NAME/KEY: misc_feature
; LOCATION: (1)...(657)
; OTHER INFORMATION: n = A,T,C or G
US-09-922-480-3

Query Match 70.0%; Score 460; DB 34; Length 657;
Best Local Similarity 100.0%; Pred. No. 1.5e-58;
Matches 657; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGAARYTNTYNTGGCGNTGYATHGTNTGYTGNGCNTTYGCMNGNAARMGNMNTTY 60
DB 1 ATGAARYTNTYNTGGCGNTGYATHGTNTGYTGNGCNTTYGCMNGNAARMGNMNTTY 60
QY 61 CCNTTYATHGGNGARGAYGAYAAAYGAYGGNCAVCNNTNCAVCNWSNTNAAAYATH 120
DB 61 CCNTTYATHGGNGARGAYGAYAAAYGAYGGNCAVCNNTNCAVCNWSNTNAAAYATH 120
QY 121 CCNTAYGGNATHMGNAAYTNCNCCNCCNNTNTATAYTMGNCCNGTNAAYACNGTNCN 180
DB 121 CCNTAYGGNATHMGNAAYTNCNCCNCCNNTNTATAYTMGNCCNGTNAAYACNGTNCN 180
QY 181 WSNATCCNGGNAAYACNTAYACNGAYACNGGNYTNCNWSNTAYCCNTGGATHYTACN 240
DB 181 WSNATCCNGGNAAYACNTAYACNGAYACNGGNYTNCNWSNTAYCCNTGGATHYTACN 240

Db 181 WSNAYCCNGGNAAYACNTAYACNGAYACGNGNYTNCNWSNTAYCCNTGGATHYNACN 240
QY 241 WSNCCNGGNTTYCCNTAYGTNTAYCAYATHMGNGGNTTYCCNTYNGCNACNCARYTNAAY 300
Db 241 WSNCCNGGNTTYCCNTAYGTNTAYCAYATHMGNGGNTTYCCNTYNGCNACNCARYTNAAY 300
QY 301 GTNCCNCCNTYNCNCCNMNGGNTTYCCNTYGTNCCNCCNWSNMGNNTTYTYSWSNGCN 360
Db 301 GTNCCNCCNTYNCNCCNMNGGNTTYCCNTYGTNCCNCCNWSNMGNNTTYTYSWSNGCN 360
QY 361 GCNGCNGCNGCNGCNGCNCNATHGCGNGCGARCCNGCGNGCGNGCGNGCGNCCNYTNACN 420
Db 361 GCNGCNGCNGCNGCNGCNCNATHGCGNGCGARCCNGCGNGCGNGCGNGCGNCCNYTNACN 420
QY 421 GCNACNCCNGTNGCGNGCGARCCNGCNGCGNGCGNGCGNGCGNGCGNGCGNCCNGCGN 480
Db 421 GCNACNCCNGTNGCGNGCGARCCNGCNGCGNGCGNGCGNGCGNGCGNGCGNCCNGCGN 480
QY 481 GARGCNCNGTNGCGNGCGARCCNGCNGCGNGCGNGCGNGCGNGCGNGCGNCCNGCGN 540
Db 481 GARGCNCNGTNGCGNGCGARCCNGCNGCGNGCGNGCGNGCGNGCGNGCGNCCNGCGN 540
QY 541 GARGCNCNGTNGCGNGCGARCCNGCNGCGNGCGNGCGNGCGNGCGNGCGNCCNGCGN 600
Db 541 GARGCNCNGTNGCGNGCGARCCNGCNGCGNGCGNGCGNGCGNGCGNGCGNCCNGCGN 600
QY 601 GCNAARCCNGCNGCNGCGARCCNGCAYCCNWSNCCNWSNYTNGARCGCNAAYCAR 657
Db 601 GCNAARCCNGCNGCNGCGARCCNGCAYCCNWSNCCNWSNYTNGARCGCNAAYCAR 657

RESULT 3
US-09-923-236-3
; Sequence 3, Application US/09923236
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; APPLICANT: Adler, David A.
; TITLE OF INVENTION: SECRETED SALIVARY ZSIG63 POLYPEPTIDE
; FILE REFERENCE: 97-71
; CURRENT APPLICATION NUMBER: US/09/923,236
; PRIOR FILING DATE: 2001-08-03
; PRIOR APPLICATION NUMBER: US 60/124,820
; PRIOR FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 657
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Degenerate polynucleotide sequence for zsig63
; NAME/KEY: misc.feature
; LOCATION: (1)...(657)
; OTHER INFORMATION: n = A,T,C or G
US-09-923-236-3

Query Match 70.0%; Score 460; DB 34; Length 657;
Best Local Similarity 100.0%; Pred. No. 1.5e-58;
Matches 657; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGAARYTNYTNTGGCGTGTATHTGTGTGTCNTTYGCGNMGNAAARMGNMNTTY 60
Db 1 ATGAARYTNYTNTGGCGTGTATHTGTGTGTCNTTYGCGNMGNAAARMGNMNTTY 60
QY 61 CCNTYATHGCGNGARGAYGAYAAAYGAYGCGNCCNYTNCAYCCNWSNYTNAAYATH 120
Db 61 CCNTYATHGCGNGARGAYGAYAAAYGAYGCGNCCNYTNCAYCCNWSNYTNAAYATH 120
QY 121 CCNTAYGGNATHMGNAAYTTCNCCNCCNYTNTAYTAYMGNCCNGTNAAYACNGTNCN 180
Db 121 CCNTAYGGNATHMGNAAYTTCNCCNCCNYTNTAYTAYMGNCCNGTNAAYACNGTNCN 180
QY 181 WSNAYCCNGGNAAYACNTAYACNGAYACGNGNYTNCNWSNTAYCCNTGGATHYNACN 240

Db 181 WSNAYCCNGGNAAYACNTAYACNGAYACGNGNYTNCNWSNTAYCCNTGGATHYNACN 240
QY 241 WSNCCNGGNTTYCCNTAYGTNTAYCAYATHMGNGGNTTYCCNTYNGCNACNCARYTNAAY 300
Db 241 WSNCCNGGNTTYCCNTAYGTNTAYCAYATHMGNGGNTTYCCNTYNGCNACNCARYTNAAY 300
QY 301 GTNCCNCCNTYNCNCCNMNGGNTTYCCNTYGTNCCNCCNWSNMGNNTTYTYSWSNGCN 360
Db 301 GTNCCNCCNTYNCNCCNMNGGNTTYCCNTYGTNCCNCCNWSNMGNNTTYTYSWSNGCN 360
QY 361 GCNGCNGCNGCNGCNGCNCNATHGCGNGCGARCCNGCGNGCGNGCGNCCNYTNACN 420
Db 361 GCNGCNGCNGCNGCNGCNCNATHGCGNGCGARCCNGCGNGCGNGCGNCCNYTNACN 420
QY 421 GCNACNCCNGTNGCGNGCGARCCNGCNGCGNGCGNGCGNGCGNGCGNCCNGCGN 480
Db 421 GCNACNCCNGTNGCGNGCGARCCNGCNGCGNGCGNGCGNGCGNGCGNCCNGCGN 480
QY 481 GARGCNCNGTNGCGNGCGARCCNGCNGCGNGCGNGCGNGCGNGCGNCCNGCGN 540
Db 481 GARGCNCNGTNGCGNGCGARCCNGCNGCGNGCGNGCGNGCGNGCGNCCNGCGN 540
QY 541 GARGCNCNGTNGCGNGCGARCCNGCNGCGNGCGNGCGNGCGNGCGNCCNGCGN 600
Db 541 GARGCNCNGTNGCGNGCGARCCNGCNGCGNGCGNGCGNGCGNGCGNCCNGCGN 600
QY 601 GCNAARCCNGCNGCNGCGARCCNGCAYCCNWSNCCNWSNYTNGARCGCNAAYCAR 657
Db 601 GCNAARCCNGCNGCNGCGARCCNGCAYCCNWSNCCNWSNYTNGARCGCNAAYCAR 657

RESULT 4
PCT-US02-08123-144
; Sequence 144, Application PC/TUS0208123
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc.
; TITLE OF INVENTION: Human Secreted Proteins
; FILE REFERENCE: PS904PCT
; CURRENT APPLICATION NUMBER: PCT/US02/08123
; CURRENT FILING DATE: 2002-03-19
; PRIOR APPLICATION NUMBER: US 60/331,287
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/306,171
; PRIOR FILING DATE: 2001-07-19
; PRIOR APPLICATION NUMBER: US 60/277,340
; PRIOR FILING DATE: 2001-03-21
; NUMBER OF SEQ ID NOS: 2048
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 144
; LENGTH: 961
; TYPE: DNA
; ORGANISM: Homo sapiens
PCT-US02-08123-144
Query Match 70.0%; Score 460; DB 1; Length 961;
Best Local Similarity 61.0%; Pred. No. 1.5e-58;
Matches 401; Conservative 103; Mismatches 153; Indels 0; Gaps 0;
QY 1 ATGAARYTNYTNTGGCGTGTATHTGTGTGTCNTTYGCGNMGNAAARMGNMNTTY 60
Db 70 ATGAAGCTTCTCTTTGGGCTGCATGTATGTGTGTTGCTTTTGCAAGGAGAGACGGTTC 129
QY 61 CCNTYATHGCGNGARGAYGAYAAAYGAYGCGNCCNYTNCAYCCNWSNYTNAAYATH 120
Db 130 CCCTTCATGTGGTAGGATGACATGACGATGGTCACCCACTTCATCCATCTCTGATATT 189
QY 121 CCNTAYGGNATHMGNAAYTTCNCCNCCNYTNTAYTAYMGNCCNGTNAAYACNGTNCN 180
Db 190 CCCTATGGATACGGAATTACCCACTCTCTTATTATCGGCCAGTGAATACAGTCCCC 249
QY 181 WSNAYCCNGGNAAYACNTAYACNGAYACGNGNYTNCNWSNTAYCCNTGGATHYNACN 240

Db 250 AGTTACCCCTGGGAATACCTACACTGACACAGGGTTACCTTCGTATCCCTCGGATTCTAACT 309
QY 241 WSNCCNGGNTTYCCNTAYGTNTAYCAYATHMGNGGNTTYCCNYTNGCNACNCARYTNAAY 300
Db 310 TCTCCTGGATTCCTCATATGCTATACATCCGTTGTTTCCCTTAGCTACTCAGTTGAAT 369
QY 301 GTNCCNCCNTYCCNCCNMNGGNTTYCCNTTYGTNCCNCCNWSNMGNNTTYTYSWNGCN 360
Db 370 GTTCTCCTCTCCCTCCCTAGGGGTTTCCCGTTGTCCCTCTTCAAGGTTTTTTTCAGCA 429
QY 361 GCGNGCNGCNGCNGCNGCNCNATHGNGCNGARCCNGCNGCNGCNGCNGCNYTNACN 420
Db 430 GCTGCAGCACCCGCTGCCACCCTATTGACGCTGAGCCTGCTCAGCTGCACCTCTTACA 489
QY 421 GGNACNCCNGTNGCNGCNGARCCNCCNGCNGCNGCNGCNGCNGCNGCNGCNGCN 480
Db 490 GCCACACCTGTAGCAGCTGAGCCTGCTGCAGAGGACCTGTTGCAGCTGAGCCTGCTGCA 549
QY 481 GARGCNCNCTNGGNGCNGARCCNCCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCN 540
Db 550 GAGGCACCTGTTGGAGCTGAGCCTGCTGCAGAGGACCTGTTGCAGCTGAGCCTGCTGCA 609
QY 541 GARGCNCNCTNGGNGCNGARCCNCCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCN 600
Db 610 GAGGCACCTGTTGGAGTGGAGCCAGCTGCAGAGGAACTTCACCACTGAGCCTGCTTACA 669
QY 601 GCNAARCCNCCNCCNGARCCNCAYCCNWSNCCNWSNYTNGARCCNCAAYCAR 657
Db 670 GCCAAGCTGCTGCCCGAGAACCTCACCTTCTCCTCTCTTGAACAGGCAATCAG 726

RESULT 5
PCT-US02-08277-108
; Sequence 108, Application PC/TUS0208277
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc.
; TITLE OF INVENTION: Human Secreted Proteins
; FILE REFERENCE: PS907PCT
; CURRENT APPLICATION NUMBER: PCT/US02/08277
; PRIOR FILING DATE: 2002-03-19
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/331,287
; PRIOR FILING DATE: 2001-07-19
; PRIOR APPLICATION NUMBER: US 60/277,340
; PRIOR FILING DATE: 2001-03-21
; NUMBER OF SEQ ID NOS: 1357
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 108
; LENGTH: 961
; TYPE: DNA
; ORGANISM: Homo sapiens
PCT-US02-08277-108

Query Match 70.0%; Score 450; DB 1; Length 961;
Best Local Similarity 61.0%; Pred. No. 1.5e-58;
Matches 401; Conservative 103; Mismatches 153; Indels 0; Gaps 0;

QY 1 ATGAARYTNTYNTGCGCNTGYATHGNTGYGTNGCNYTNGCNMGNMGNNTTY 60
Db 70 ATGAAGCTTCTCCTTTGGGCCCTGCATTTGATGTGTTGCTTTTGAAGGAAGAGCGGTTTC 129
QY 61 CCNTTYATHGGNGARGAYGAYAYGAYGAYGAYCCNYTNCAYCCNWSNYTNAAYATH 120
Db 130 CCCTTCATTGGTGAGGATGACATGACATGCTGACCACTTCATCATCTCTGATATT 189
QY 121 CCNTAYGGNATHMGNAAYTNCNCCNCCNYTNTAYTAYMGNCNCTNGTNAAYACNGTNCN 180
Db 190 CCTTATGGCATACGGAATTTACACCTCCTCTTATTATCGCCCACTGTAATACAGTCCCC 249
QY 181 WSNAYCCNGGNAAYACNTAYACNGAYACNGNYTNCNWSNTAYCCNTGGATHVHTNACN 240
Db 250 AGTTACCCCTGGGAATACCTACACTGACACAGGGTTACCTTCGTATCCCTCGGATTCTAACT 309

QY 241 WSNCCNGGNTTYCCNTAYGTNTAYCAYATHMGNGGNTTYCCNYTNGCNACNCARYTNAAY 300
Db 310 TCTCCTGGATTCCTCATATGCTATACATCCGTTGTTTCCCTTAGCTACTCAGTTGAAT 369
QY 301 GTNCCNCCNTYCCNCCNMNGGNTTYCCNTTYGTNCCNCCNWSNMGNNTTYTYSWNGCN 360
Db 370 GTTCTCCTCTCCCTCCCTAGGGGTTTCCCGTTGTCCCTCTTCAAGGTTTTTTTCAGCA 429
QY 361 GCGNGCNGCNGCNGCNGCNCNATHGNGCNGARCCNGCNGCNGCNGCNGCNYTNACN 420
Db 430 GCTGCAGCACCCGCTGCCACCCTATTGACGCTGAGCCTGCTCAGCTGCACCTCTTACA 489
QY 421 GGNACNCCNGTNGCNGCNGARCCNCCNGCNGCNGCNGCNGCNGCNGCNGCNGCN 480
Db 490 GCCACACCTGTAGCAGCTGAGCCTGCTGCAGAGGACCTGTTGCAGCTGAGCCTGCTGCA 549
QY 481 GARGCNCNCTNGGNGCNGARCCNCCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCN 540
Db 550 GAGGCACCTGTTGGAGCTGAGCCTGCTGCAGAGGACCTGTTGCAGCTGAGCCTGCTGCA 609
QY 541 GARGCNCNCTNGGNGCNGARCCNCCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCN 600
Db 610 GAGGCACCTGTTGGAGTGGAGCCAGCTGCAGAGGAACTTCACCACTGAGCCTGCTTACA 669
QY 601 GCNAARCCNCCNCCNGARCCNCAYCCNWSNCCNWSNYTNGARCCNCAAYCAR 657
Db 670 GCCAAGCTGCTGCCCGAGAACCTCACCTTCTCCTCTCTTGAACAGGCAATCAG 726

RESULT 6
PCT-US02-08278-143
; Sequence 143, Application PC/TUS0208278
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc.
; TITLE OF INVENTION: Human Secreted Proteins
; FILE REFERENCE: PS902PCT
; CURRENT APPLICATION NUMBER: PCT/US02/08278
; PRIOR FILING DATE: 2002-03-19
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/331,287
; PRIOR FILING DATE: 2001-07-19
; PRIOR APPLICATION NUMBER: US 60/306,171
; PRIOR FILING DATE: 2001-03-21
; NUMBER OF SEQ ID NOS: 1988
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 143
; LENGTH: 961
; TYPE: DNA
; ORGANISM: Homo sapiens
PCT-US02-08278-143

Query Match 70.0%; Score 460; DB 1; Length 961;
Best Local Similarity 61.0%; Pred. No. 1.5e-58;
Matches 401; Conservative 103; Mismatches 153; Indels 0; Gaps 0;

QY 1 ATGAARYTNTYNTGCGCNTGYATHGNTGYGTNGCNYTNGCNMGNMGNNTTY 60
Db 70 ATGAAGCTTCTCCTTTGGGCCCTGCATTTGATGTGTTGCTTTTGAAGGAAGAGCGGTTTC 129
QY 61 CCNTTYATHGGNGARGAYGAYAYGAYGAYGAYCCNYTNCAYCCNWSNYTNAAYATH 120
Db 130 CCCTTCATTGGTGAGGATGACATGACATGCTGACCACTTCATCATCTCTGATATT 189
QY 121 CCNTAYGGNATHMGNAAYTNCNCCNCCNYTNTAYTAYMGNCNCTNGTNAAYACNGTNCN 180
Db 190 CCTTATGGCATACGGAATTTACCACTCCTCTTATTATCGCCCACTGTAATACAGTCCCC 249
QY 181 WSNAYCCNGGNAAYACNTAYACNGAYACNGNYTNCNWSNTAYCCNTGGATHVHTNACN 240
Db 250 AGTTACCCCTGGGAATACCTACACTGACACAGGGTTACCTTCGTATCCCTCGGATTCTAACT 309


```
QY 61 CCNTYATHGGNGARGAYGAYAAAYGAYGAGGNCAYCCNYTNCAYCCNWSNYTNAAYATH 120
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 188 CCCTTCAUTGGTGAGGATGACAATGACCATGTCACCCACACTTCATCCATCTCTGAATATT 247

QY 121 CCNTAYGGNATHMGNAAYTNCNCNCNCNTNTAYTAYMGNCNGTNAAYACNGTNCN 180
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 248 CCTTATGGCATACGGAATTTACCACTCCTCTTTATTATCGCCAGTGAATACAGTCCCC 307

QY 181 WSNATYCCNGGNAAYACNTAYACNGAYACNGNYTNCNWSNTAYCCNTGGGATHYTACN 240
::|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 308 AGTTACCTCGGAATACCTTACACTGACACAGGTTACCTTCGTATCCCTGGATTTCTAACT 367

QY 241 WSNCCNGNTTYCCNTAYGTNTAYCAYATHMGNGNTTYCCNTYNGCNACNCARYTNAAY 300
::|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 368 TCTCCTGGATTCCCTATGCTATACATCCCGTGGTTTCCCTTAGCTACTCAGTTGAAT 427

QY 301 GTNCCNCNTYTNCCNMGNGNTTYCCNTYGTNCCNWSNMNGNTTYTWTWSNGCN 360
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 428 GTTCTCCTCTCCCTAGGGGTTTCCCGTTTGTCCCTCTCAAGGTTTTTTTCAGCA 487

QY 361 GCGNGCNCNGCNCNCNCNCNCNCNCNCNCNCNCNCNCNCNCNCNCNCNCNCNCNCN 420
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 488 GCTGCAGCACCCGCTGCCCCACCTATTGCAGCTGAGCCTGCTGCAGCTGCACCTCTTACA 547

QY 421 GCNACNCNGTNGCNGARCCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCN 480
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 548 GCCACACCTGTAGCAGCTGAGCCTGCTGCAGGGGCCCTGTTGCAGCTGAGCCTGCTGCA 607

QY 481 GARGCNCNGTNGGNGCNGARCCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCN 540
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 608 GAGGCACCTGTTGAGCTGAGCCTGCTGCAGAGGACCTGTTGCAGCTGAGCCTGCTGCA 667

QY 541 GARGCNCNGTNGGNGCNGARCCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCN 600
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 668 GAGGCACCTGTTGAGTGGAGCCAGCTGCAGAGGAACCTTCACCACTGAGCCTGCTTACA 727

QY 601 GCNAARCCNGCNCNCNCNCNCNCNCNCNCNCNCNCNCNCNCNCNCNCNCNCNCNCN 657
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 728 GCCAAGCCTGCTGCCCAAGAACCTCACCCCTTCTCCCTCTCTTGAACAGGCAATACAG 784

RESULT 11
US-09-922-480-1
; Sequence 1, Application US/09922480
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; APPLICANT: Adler, David A.
; TITLE OF INVENTION: SECRETED SALIVARY ZSIG63 POLYPEPTIDE
; FILE REFERENCE: 97-71
; CURRENT APPLICATION NUMBER: US/09/922,480
; CURRENT FILING DATE: 2001-08-03
; PRIOR APPLICATION NUMBER: US 60/124,820
; PRIOR FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 1008
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (128)...(784)
US-09-922-480-1

Query Match 70.0%; Score 460; DB 34; Length 1008;
Best Local Similarity 61.0%; Pred. No. 1.5e-58;
Matches 401; Conservative 103; Mismatches 153; Indels 0; Gaps 0;

QY 1 ATGAARTYNTYNTGGGNTGYATHGNTGYTGNCNTTYGNCNMGNAARMNGNTTY 60
|||||:|:|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 128 ATGAAGCTTCTCCTTTGGGCTGCAITGTATGTGTGTTTGTCTTTTGAAGGAAGACGGTTC 187

QY 61 CCNTYATHGGNGARGAYGAYAAAYGAYGAGGNCAYCCNYTNCAYCCNWSNYTNAAYATH 120
```

```
Db 188 CCCTTCAUTGGTGAGGATGACAATGACCATGTCACCCACACTTCATCCATCTCTGAATATT 247

QY 121 CCNTAYGGNATHMGNAAYTNCNCNCNCNTNTAYTAYMGNCNGTNAAYACNGTNCN 180
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 248 CCTTATGGCATACGGAATTTACCACTCCTCTTTATTATCGCCAGTGAATACAGTCCCC 307

QY 181 WSNATYCCNGGNAAYACNTAYACNGAYACNGNYTNCNWSNTAYCCNTGGGATHYTACN 240
::|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 308 AGTTACCTCGGAATACCTTACACTGACACAGGTTACCTTCGTATCCCTGGATTTCTAACT 367

QY 241 WSNCCNGNTTYCCNTAYGTNTAYCAYATHMGNGNTTYCCNTYNGCNACNCARYTNAAY 300
::|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 368 TCTCCTGGATTCCCTATGCTATACATCCCGTGGTTTCCCTTAGCTACTCAGTTGAAT 427

QY 301 GTNCCNCNTYTNCCNMGNGNTTYCCNTYGTNCCNWSNMNGNTTYTWTWSNGCN 360
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 428 GTTCTCCTCTCCCTAGGGGTTTCCCGTTTGTCCCTCTCAAGGTTTTTTTCAGCA 487

QY 361 GCGNGCNCNGCNCNCNCNCNCNCNCNCNCNCNCNCNCNCNCNCNCNCNCNCNCN 420
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 488 GCTGCAGCACCCGCTGCCCCACCTATTGCAGCTGAGCCTGCTGCAGCTGCACCTCTTACA 547

QY 421 GCNACNCNGTNGCNGARCCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCN 480
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 548 GCCACACCTGTAGCAGCTGAGCCTGCTGCAGGGGCCCTGTTGCAGCTGAGCCTGCTGCA 607

QY 481 GARGCNCNGTNGGNGCNGARCCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCN 540
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 608 GAGGCACCTGTTGAGCTGAGCCTGCTGCAGAGGACCTTCACCACTGAGCCTGCTTACA 667

QY 541 GARGCNCNGTNGGNGCNGARCCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCN 600
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 668 GAGGCACCTGTTGAGTGGAGCCAGCTGCAGAGGAACCTTCACCACTGAGCCTGCTTACA 727

QY 601 GCNAARCCNGCNCNCNCNCNCNCNCNCNCNCNCNCNCNCNCNCNCNCNCNCNCN 657
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 728 GCCAAGCCTGCTGCCCAAGAACCTCACCCCTTCTCCCTCTCTTGAACAGGCAATACAG 784

RESULT 12
US-09-923-236-1
; Sequence 1, Application US/09923236
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; APPLICANT: Adler, David A.
; TITLE OF INVENTION: SECRETED SALIVARY ZSIG63 POLYPEPTIDE
; FILE REFERENCE: 97-71
; CURRENT APPLICATION NUMBER: US/09/923,236
; CURRENT FILING DATE: 2001-08-03
; PRIOR APPLICATION NUMBER: US 60/124,820
; PRIOR FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 1008
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (128)...(784)
US-09-923-236-1

Query Match 70.0%; Score 460; DB 34; Length 1008;
Best Local Similarity 61.0%; Pred. No. 1.5e-58;
Matches 401; Conservative 103; Mismatches 153; Indels 0; Gaps 0;

QY 1 ATGAARTYNTYNTGGGNTGYATHGNTGYTGNCNTTYGNCNMGNAARMNGNTTY 60
|||||:|:|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 128 ATGAAGCTTCTCCTTTGGGCTGCAITGTATGTGTGTTTGTCTTTTGAAGGAAGACGGTTC 187

QY 61 CCNTYATHGGNGARGAYGAYAAAYGAYGAGGNCAYCCNYTNCAYCCNWSNYTNAAYATH 120
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

Db 188 CCCTTCATTGGTGGAGATGACAATGACGATGGTCACCCACTTCATCCACTCTCTGAATATT 247
Qy 121 CNTAYGNGNATHMGNAAYTTCNCCNCNNTNTAYTAYMNCNGNTNAYACNGTNCN 180
Db 248 COTATGACATACGGAATTTACCACTCTCTTTATTATGCCCGAGTGAATACAGTCCCC 307
Qy 181 WNTAYCCNGGNAAYACNTAYACGAYACNGGNTNCCNWSNTAYCNGTGGATHYTNA 240
Db 308 AGTTACCTGGGATCTACTACACTGACACAGGTTACCTTCGTATCCCTGGATTCTA 367
Qy 241 WSNCCNGNTTCCNTAYGTNTAYATYATHMGNGNTTYCCNTYNGNACNARYTNA 300
Db 368 TCTCCTGGATTCCTCTATGTCTATCATCCGTGGTTTTCCTTTAGCTACTCAGTTGA 427
Qy 301 GTNCCNCCNTNCCNMGNGNTTYCCNTYGTNCCNCCNWSNMGTNTTYWSNGCN 360
Db 428 GTTCTCTCTCTCCCTCCCTAGGGGTTTCCCGTTTGTCCCTCTCAAGGTTTTTTTC 487
Qy 361 GCNGCNGCNGCNGCNGCNGCNCNATHGCGNGCGARCCNGCNGCNGCNGCNGCNGCN 420
Db 488 GCTGACGACCCGCTGCCACCTATTTCAGCTGAGCCTGCTGACGCTGCACCTTTACA 547
Qy 421 GCNACNCGNTNGCNGCNGARCCNCCNGCNGCNGCNGCNGCNGCNGCNGCNGCN 480
Db 548 GCCACACCTGTAGCAGCTGAGCCTGCTGCAGGGGCCCTGTTGCAGCTGAGCCTGCTG 607
Qy 481 GARGCNCNGTNGCNGCNGARCCNCCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCN 540
Db 608 GAGGCACCTGTTGGAGCTGAGCCTGCTGCAGAGGACCTGTTGCAGCTGAGCCTGCTG 667
Qy 541 GARGCNCNGTNGCNGCNGARCCNCCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCN 600
Db 668 GAGGCACCTGTTGGAGTGAGCAGCTGCAGAGGAACCTTCACACAGCTGAGCCTGCTA 727
Qy 601 GCNARCCNCGCNGCNGARCCNCCNAYCCNWSNCCNWSNTYNGARCCNAYCAR 657
Db 728 GCCAAGCCTGTGCCCGAGAACCTCACCTCTCTCCCTCTCTTGAACAGGCAATCAG 784

RESULT 13
PCT-US98-13813-9
; Sequence 9, Application PC/TUS9813813
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John M.
; APPLICANT: LaVallie, Edward R.
; APPLICANT: Racie, Lisa A.
; APPLICANT: Treacy, Maurice
; APPLICANT: Spaulding, Vikki
; APPLICANT: Agostino, Michael J.
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: MA
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US98/13813
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Sprunger, Suzanne A.
; REGISTRATION NUMBER: 41.323
; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (617) 498-8284
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1325 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
PCT-US98-13813-9

Query Match 69.5%; Score 456.8; DB 1; Length 1325;
Best Local Similarity 60.7%; Pred. No. 4.6e-58;
Matches 399; Conservative 103; Mismatches 155; Indels 0; Gaps 0;

Qy 1 ATCAARTYNTYNTGCGCTGATGNTGTYGTNGCNTTYGCNMGNAARMGNMNTTY 60
Db 127 ATGAAGCTTCTCCTTTGGGCTCGATTGATGTGTGCTTTTTCGAAGGAGAGACGGTTC 186
Qy 61 CCNTTYATHGGNGARGAYGAYAYGAYGAYGNCAYCCNYTNCAYCCNWSNTYNAAYATH 120
Db 187 CCCTTCATTGGTGGAGTGAATGACATGAGTGGTCAACCCACTTCATCCATCTCTGAATA 246
Qy 121 CCNTAYGNGNATHMGNAAYTTCNCCNCNNTNTAYTAYMNCNGNTNAYACNGTNCN 180
Db 247 CCTATGGCATACGGAATTTACCACTCTCTTTATTATCGCCCGAGTGAATACAGTCCC 306
Qy 181 WNTAYCCNGGNAAYACNTAYACGAYACNGGNTYTCNWSNTAYCNGTGGATHYTNA 240
Db 307 AGTTACCTGGGATCTACTACACTGACACAGGTTTACCTTCGTATCCCTGGATTCTA 366
Qy 241 WSNCCNGNTTYCCNTAYGTNTAYATYATHMGNGNTTYCCNTYNGCNGCNGCNGCN 300
Db 367 TCTCCTGGATTCCCTATGCTATCATCCGTGGTTTTCCTTTAGCTACTCAGTTGAAT 426
Qy 301 GTNCCNCCNTNCCNMGNGNTTYCCNTYGTNCCNCCNWSNMGTNTTYWSNGCN 360
Db 427 GTTCTCTCTCTCCCTCTAGGGGTTTCCCGTTTGTCCCTCTCAAGGTTTTTTTC 486
Qy 361 GCNGCNGCNGCNGCNGCNGCNCNATHGCGNGCGARCCNGCNGCNGCNGCNGCNGCN 420
Db 487 GCTGACGACCCGCTGCCACCTATTTCAGCTGAGCCTGCTGCAGCTGCACCTTTACA 546
Qy 421 GCNACNCGNTNGCNGCNGARCCNCCNGCNGCNGCNGCNGCNGCNGCNGCNGCN 480
Db 547 TCCACACCTGTAGCATCTGAGCCTGCTGCAGGGGCCCTGTTGCAGCTGAGCCTGCTGA 606
Qy 481 GARGCNCNGTNGCNGCNGARCCNCCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCN 540
Db 607 GAGGCACCTGTTGGAGCTGAGCCTGCTGCAGAGGACCTGTTGCAGCTGAGCCTGCTGA 666
Qy 541 GARGCNCNGTNGGNGTNGARCCNCCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCN 600
Db 667 GAGGCACCTGTTGGAGTGAGCAGCTGCAGAGGAACCTTCACACAGCTGAGCCTGCTA 726
Qy 601 GCNARCCNCGCNGCNGARCCNCCNAYCCNWSNCCNWSNTYNGARCCNAYCAR 657
Db 727 GCCAAGCCTGTGCCCGAGAACCTCACCTCTCTCCCTCTCTTGAACAGGCAATCAG 783

RESULT 14
US-08-958-304-9
; Sequence 9, Application US/08958304
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John M.
; APPLICANT: LaVallie, Edward R.
; APPLICANT: Racie, Lisa A.
; APPLICANT: Treacy, Maurice
; APPLICANT: Spaulding, Vikki
; APPLICANT: Agostino, Michael J.
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
; TITLE OF INVENTION: ENCODING THEM


```
Db 528 GCAGCAGCAGCAGATCTCTCGGCGCAGCAGCAGCAACAGCAGCAGCAGCAGCAGCA 587
Qy 517 CNGTNGCNGCNGARCCNGCNGCNGARCCNGCNGTNGTNGARCCNGCNGCNGAR 576
Db 588 GCAACAGCAACAGCAGCAGCAGCAACAGCAACAGCAACAGCAGCAGCAGCAGCA 647
Qy 577 CCNWSNCCNGCNGARCCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGC 636
Db 648 ACAGGGGGCTCTCCCAACCCAGCCCGCAGTCCCGAGCCCGAGTTCACAGCGCGGGCT 707
Qy 637 WSNYTINGARCARGC 650
Db 708 TCAGCAGACCCAGC 721

RESULT 9
US-09-724-676A-2614
; Sequence 2614, Application US/09724676A
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724.676A
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2614
; LENGTH: 1051
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-724-676A-2614

Query Match 12.4%; Score 81.2; DB 5; Length 1051;
Best Local Similarity 38.2%; Pred. No. 2.7e-05;
Matches 120; Conservative 23; Mismatches 171; Indels 0; Gaps 0;

Qy 337 CCNCCNWSNMGNNTTYYWSNGCNGCNGCNGCNGCNGCNGCNGCNGCNGAR 396
Db 408 CCGTTCAACAGCCATCTCTAGCAGCAGCAGCAGCAACAGCAACAGCAACAGCA 467
Qy 397 CCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNG 456
Db 468 GCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 527
Qy 457 CCNGTNGCNGCNGARCCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNG 516
Db 528 GCAGCAGCAGCAGATCTCTCGGCGCAGCAGCAGCAACAGCAACAGCAGCAGCAG 587
Qy 517 CNGTNGCNGCNGARCCNGCNGCNGCNGCNGCNGTNGTNGARCCNGCNGCNGAR 576
Db 588 GCAACAGCAACAGCAGCAGCAGCAGCAACAGCAACAGCAACAGCAGCAGCAGCA 647
Qy 577 CCNWSNCCNGCNGARCCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGC 636
Db 648 ACAGGGGGCTCTCCCAACCCAGCCCGCAGTCCCGAGCCCGAGTTCACAGCGCGGGCT 707
Qy 637 WSNYTINGARCARGC 650
Db 708 TCAGCAGACCCAGC 721

RESULT 10
US-09-724-676-2612
; Sequence 2612, Application US/09724676
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724.676
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2612
```

```
; LENGTH: 2280
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-724-676-2612

Query Match 12.4%; Score 81.2; DB 5; Length 2280;
Best Local Similarity 38.2%; Pred. No. 3.2e-05;
Matches 120; Conservative 23; Mismatches 171; Indels 0; Gaps 0;

Qy 337 CCNCCNWSNMGNNTTYYWSNGCNGCNGCNGCNGCNGCNGCNGCNGCNGAR 396
Db 1637 CCGTTCAACAGCCATCTCTAGCAGCAGCAGCAGCAACAGCAACAGCAACAG 1696
Qy 397 CCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNG 456
Db 1697 GCAGCAGCAGCAGCAACAGCAACAGCAGCAGCAGCAGCAGCAGCAGCAGCA 1756
Qy 457 CCNGTNGCNGCNGARCCNGCNGCNGCNGCNGCNGTNGGNGCNGCNGCNGCNG 516
Db 1757 GCAGCAGCAGCAGATCTCTCGGCGCAGCAGCAGCAACAGCAACAGCAGCAG 1816
Qy 517 CCNGTNGCNGCNGARCCNGCNGCNGCNGCNGCNGCNGTNGGNGCNGCNGCNG 576
Db 1817 GCAACAGCAACAGCAGCAGCAGCAGCAACAGCAACAGCAACAGCAGCAGCA 1876
Qy 577 CCNWSNCCNGCNGARCCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNG 636
Db 1877 ACAGGGGGCTCTCCCAACCCAGCCCGCAGTCCCGAGCCCGAGTTCACAGCG 1936
Qy 637 WSNYTINGARCARGC 650
Db 1937 TCAGCAGACCCAGC 1950

RESULT 11
US-09-724-676A-2612
; Sequence 2612, Application US/09724676A
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724.676A
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2612
; LENGTH: 2280
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-724-676A-2612

Query Match 12.4%; Score 81.2; DB 5; Length 2280;
Best Local Similarity 38.2%; Pred. No. 3.2e-05;
Matches 120; Conservative 23; Mismatches 171; Indels 0; Gaps 0;

Qy 337 CCNCCNWSNMGNNTTYYWSNGCNGCNGCNGCNGCNGCNGCNGCNGCNGAR 396
Db 1637 CCGTTCAACAGCCATCTCTAGCAGCAGCAGCAGCAACAGCAACAGCAACAG 1696
Qy 397 CCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNG 456
Db 1697 GCAGCAGCAGCAGCAACAGCAACAGCAGCAGCAGCAGCAGCAGCAGCAGCA 1756
Qy 457 CCNGTNGCNGCNGARCCNGCNGCNGCNGCNGCNGTNGGNGCNGCNGCNGCNG 516
Db 1757 GCAGCAGCAGCAGATCTCTCGGCGCAGCAGCAGCAACAGCAACAGCAGCAG 1816
Qy 517 CCNGTNGCNGCNGARCCNGCNGCNGCNGCNGCNGTNGGNGCNGCNGCNGCNG 576
Db 1817 GCAACAGCAACAGCAGCAGCAGCAGCAACAGCAACAGCAACAGCAGCAGCA 1876
Qy 577 CCNWSNCCNGCNGARCCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNG 636
```


[illegible]

```

RESULT 13
US-09-724-676A-2621
; Sequence 2621, Application US/09724676A
; GENERAL INFORMATION:
; APPLICANT: CompuGen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 CompuGen
; CURRENT APPLICATION NUMBER: US/09/724, 676A
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2621
; LENGTH: 6723
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-724-676A-2621

```

Query Match	12.4%;	Score 81.2;	DB 5;	Length 6723;
Best Local Similarity	38.2%;	Pred. No. 4.1e-05;		
Matches 120; Conservative	23;	Mismatches 171;	Indels 0;	Gaps 0;

Qy 337 CCNCNWSNMGNNTYTTYWSNGCNGCNGCNGCNGCCNCCNATGCGCNGAR 396
|| : : : || || || || || || : || :
Db 6080 CCGTTTCAACAGCCATCTACTGAGCAGCAGCAGCAACAGCAGCAACAACGA 6139
Qy 397 CCNGCNGCNGCNGCNCNYTNACNGCNAACNGCNGCNGCNGCNGCNGCNGCN 456
Dbb 6140 GCAGCAGCAGCAGCAACAGCAGCAGCAGCAGCAGCAGCATACCACATCCGGCAGCA 6199
Qy 457 CCNGTCNGCNGCNGCNGCNGCNGARGCNCNCNGTGNGCNGCNGARCCNGCNGCNGARGCN 516
Db 6200 GCAGCAGCAGCATCTCGGCAGCAGCAGCAACAGCAGCAGCAGCAGCAGCAGCA 6259
Qy 517 CCNGTCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGTGNGTGNGCNGCNGCNGAR 576
Db 6260 GCAACAGCAACAGCAGCAGCAACAGCAACAGCAACAGCAACAGCAGCAACAGCAGCA 6319
Qy 577 CCNWSNCCNGCNGCNGCNGCNGCNGCNGARCCNGCNGCNGCNGCNGCNCATCCNWSNCCN 636
Db 6320 ACAGCGGGCTCTCCCAACCCAGCCCCAGTCCAGCGCCCAGTTCCAGCGCCAGGGGT 6379
Qy 637 WSNYTNGARCARGC 650
Db 6380 TCAGCACACCCAGC 6393

RESULT 14
US-10-240-425-1294
; Sequence 1294, Application US/10240425
; GENERAL INFORMATION:
; APPLICANT: Williams, Amanda
; APPLICANT: Boland, Joseph F.
; APPLICANT: Lord, Reginald V.
; APPLICANT: Alvarez, Chris
; APPLICANT: Wetzel, Jon C.
; APPLICANT: Scherf, Uwe
; APPLICANT: Vockley, Joseph G.
; TITLE OF INVENTION: Gene Expression Profiles in Esophageal Tissue
; FILE REFERENCE: 44921-5026
; CURRENT APPLICATION NUMBER: US/10/240,425
; CURRENT FILING DATE: 2002-09-30
; PRIORITY APPLICATION NUMBER: PCT/US01/09847
; PRIOR FILING DATE: 2001-03-28
; PRIOR APPLICATION NUMBER: US 60/193,446
; PRIOR FILING DATE: 2000-03-31
; NUMBER OF SEQ ID NOS: 1588
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1294
; LENGTH: 2108
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. M13903
US-10-240-425-1294

Query Match	12.1%	Score 79.6;	DB 6;	Length 2108;
Best Local Similarity	41.7%;	Pred. No. 5.8e-05;		
Matches 110; Conservative 15; Mismatches 139;			Indels 0;	Gaps
Qy	356	SNGCNGCNGCNGCNGCNGCNGCNCNCNCNATGCGCNGCNGARGCNGCNGCNGCNCNCNY	415	
Db	843	CAGCTGGAGGTCCAGAGGAGCAGATGGGGCAGCTGAAGTACCTTGGAAACAGCAGAGGGG	902	
Qy	416	TNACNGCNAACNCNGTNGCNGCNGARFCNGCNGCNGGNGCNCNCCTNGCNGCNGARCCNG	475	
Db	903	CAGCTGAAGCACCTGGATCAGCAGGAGAGCAGCAGCAGCTCCAGACAGCAGATGGGG	962	
Qy	476	CNGCNGARGCNCNGTNGCNGCNGARCCNGCNGCNGCNGARCCNGTNGCNGCNGARCCNG	535	
Db	963	CAGCTGAAGCACCTGGAGCAGCAGGAGGGGCGACCTAAGCATCTTGAGCAGCAGAGGGG	1022	
Qy	536	CNGCNGARGCNCNCNGTNGGNGTNGARCCNGCNGCNGCNGARGARCCNMNSCNCNGCNGARCCNG	595	

```
Db 1023 CAACTGGAGCAGCTGGAGGAGGAGGGGAGCTGAAGCACCTGGAGCAGGAGGGG 1082
QY 596 CNACNGCNAARCGNGCGNGCCNG 619
      | | | | | | | | | | | | | |
Db 1083 CAGCTGGAGCACCTGGAGCACCAG 1106

RESULT 15
US-10-144-771-43522/c
; Sequence 43522, Application US/10144771
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig
; TITLE OF INVENTION: HUMAN GENOME DISCOVERY SYSTEM AND USES THEREOF
; FILE REFERENCE: CL001321
; CURRENT APPLICATION NUMBER: US/10/144,771
; CURRENT FILING DATE: 2002-05-15
; NUMBER OF SEQ ID NOS: 47235
; SEQ ID NO 43522
; LENGTH: 792
; TYPE: DNA
; ORGANISM: HUMAN
US-10-144-771-43522

Query Match      12.1%; Score 79.2; DB 6; Length 792;
Best Local Similarity 41.2%; Pred. No. 5.5e-05;
Matches 121; Conservative 12; Mismatches 161; Indels 0; Gaps 0;

QY 356 SNGCNGCNGCNGCNGCNGCNCNCNATHGCGNGCNGARCCNGCNGCNGCNGCNGCNY 415
      : | | | | | | | | | | | | | | | | | | | | | | | |
Db 633 CTGCTGTAGGACACCTGCTGCGGCCCCAGCTGTGTGTGTCCAGCTGTCGAGACCCCA 574

QY 416 TNACNGCNACNGCNGCNGCNGARCCNGCNGCNGCNGCNGCNGARCCNG 475
      | | | | | | | | | | | | | | | | | | | | | | | |
Db 573 GTGCTCCAGTCTGTGTGCTGCCAGCCACCTGCTGTGCGCCCCAGCTGTTCATTCTAG 514

QY 476 CNGCNGARCGCNGCNGCNGCNGARCCNGCNGCNGCNGCNGCNGCNGARCCNG 535
      | | | | | | | | | | | | | | | | | | | | | | | |
Db 513 CTGCTCAGGCGCTTCCTGCTGCGGCGCTAGCTGTTCATTCTAGCTGTCAGGCGCTTC 454

QY 536 CNGCNGARCGCNGCNGCNGCNGCNGARCCNGCNGCNGCNGCNGCNGARCCNG 595
      | | | | | | | | | | | | | | | | | | | | | | | |
Db 453 TTGCTGCCGCCCGCAGCTGTGTGTGTGTCAGCTGTGTCAGACCCAGCTGTCATCTCCAG 394

QY 596 CNACNGCNAARCGNGCNGCNGCNGARCCNCAYCCNWSNCCNWSNYTINGARCA 649
      | | | | | | | | | | | | | | | | | | | | | | | |
Db 393 CTGCTGCCGCCCGCAGCTGTGCCAGACCACTGCTGCGGCCCGCAGCATGCTCTAG 340
```

Search completed: February 12, 2003, 11:53:07
Job time : 187.124 secs

	CNS06NOT	1024 bp	DNA linear	GSS 04-JUL-2001
LOCUS	T3 end of clone AU0AA015H08 of library AU0AA from strain CBS 3082			
DEFINITION	Saccharomyces kluyveri genomic survey sequence.			
ACCESSION	AL407043			
VERSION	AL407043.1	GI:12171906		
KEYWORDS	GSS			
SOURCE	Saccharomyces kluyveri.			
ORGANISM	Saccharomyces kluyveri			
REFERENCE	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;			
AUTHORS	Saccharomycetales; Saccharomycetaceae; Saccharomycetes. Souchiet,J.L., Aigle,M., Artigueave,F., Blandin,G., Bolin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S., de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Liorente,B., Malpertuy,A., Neuveglise,C., Ozier-Kalogiropoulos,O., Potier,S., Sauvin,W., Tekala,F., Toffano-Nioche,C., Weslowski-Louvel,M., Winker,P. and Weissenbach,J.			
TITLE	Genomic exploration of the hemiascomycetous yeasts: 1. A set of yeast species for molecular evolution studies			
JOURNAL	FEMS Lett. 487 (1), 3-12 (2000)			
MEDLINE	20584711			
PUBMED	11152876			
REFERENCE	2 (bases 1 to 1024) Gaillardin,C. and Casaregola,S. Neueglise,C., Bon,E., Lepingle,A., Wincker,P., Artigueave,F., Gaillardin,C. and Casaregola,S. Genomic exploration of the hemiascomycetous yeasts: 9.			
AUTHORS	Saccharomyces kluyveri FEMS Lett. 487 (1), 56-60 (2000)			
TITLE	Direct Submission			
JOURNAL	Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage, seqrefgenoscope.cns.fr - web : www.genoscope.cns.fr			
MEDLINE	This GSS is part of a random genomic sequencing program of thirteen yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii, Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila, Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 kb were prepared and both extremities were sequenced. See keywords for description of this sequence and for the sequence of the other extremity of this insert.			
PUBMED	Location/Qualifiers 1..1024 /organism="Saccharomyces kluyveri" /strain="CBS 3082"			
AUTHORS	/db_xref="taxon:4934" /clone="AU0AA015H08" /clone_lib="AU0AA" /note="end = T3"			
JOURNAL	BASE COUNT 271 a 244 c 209 g 299 t 1 others			
TITLE	ORIGIN			
COMMENT	Query Match 15.7%; Score 103.2; DB 17; Length 1024; Best local similarity 43.6%; Pred. No. 1.3e-10; Matches 127; Conservative 23; Mismatches 141; Indels 0; Gaps 0;			
	QY 333 YGTCCGCCNNMNGNTTTTYYWSNCGNCGNCNCNCNCNCNCNCNCNCNCNCNCNCNCNCGC 392			
	DB 103 TGCTTCGGCAGTAGTAATCTACTGCTGTAGAACCTGCTCTGTAGAACCCTGCTGCTGT 162			
	QY 393 NGARCCNGCNGCNGCNGCNGCNVYNACNGCNACNCNCNTGNCNGCNGCNGCNGCNGCNG 452			
	DB 163 AGAACCCTGCCTGTAGAACCTACTGCTGTAGAACCTACTGCTGTAGAACCTACTGCTGTG 222			

	BASE COUNT	30 a	195 c	148 g	153 t	ORIGIN
Query Match			15.1%	Score 99;	DB 17; Length 526;	
Best Local Similarity			43.1%;	Pred. No. 7.1e-10;		
Matches 127; Conservative 21;	Mismatches 147; Indels 0; Gaps 0;					
QY 356	SNGCNGCNGCNCNGCNCNGCNCNCNATHGCGCNGCNGARCCNCNGCNGCNGCNCNY	415				
	: : : : : : : : : : : : : : :					
Db 454	CAGCGGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG	395				
	: : : : : : : : : : : : : : :					
QY 416	TNACNGCNACNCNTGNCNGCNGARCCNGCNGCNGCNCNTGTGNCNGCARCCNG	475				
	: : : : : : : : : : : : : : :					
Db 394	CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG	335				
	: : : : : : : : : : : : : : :					
QY 476	CNGCNGARCCNCNTGGNGCNGARCCNCNGCNGCNGCNCNTGTGNCNGCARCCNG	535				
	: : : : : : : : : : : : : : :					
Db 334	CAGCTGAGACAGCAGCAGCAGCGGCAGCGCAGCAGCAGCAGCAGCAGCAGCAG	275				
	: : : : : : : : : : : : : : :					
QY 536	CNGCNGARCCNCNTGGNGTNGARCCNCNGCNGCNGARCCNCWNSCNCNGCARCCNG	595				
	: : : : : : : : : : : : : : :					
Db 274	CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG	215				
	: : : : : : : : : : : : : : :					
QY 596	CNACNGCNARCCNCNCNCNGARCCNCAYCCNWSNCNWSNYTNGARCARGC	650				
	: : : : : : : : : : : : : : :					
Db 214	CAGCAGCAGCAGCAGCAGCTGAGCAGCGGCAGCAGCAGCAGCGGCAGCAGCGGCG	160				
	: : : : : : : : : : : : : : :					

RESULT 6	CNS06PEX	1070 bp	DNA	linear	GSS 06-JUL-2001
LOCUS	CNS06PEX				
DEFINITION	T3 end of clone AV0AA014A04 of library AV0AA from strain CBS 379 of Saccharomyces exiguus, genomic survey sequence.				
ACCESSION	AL409279				
VERSION	AL409279.1	GI:12176654			
KEYWORDS	GSS.				
SOURCE	Saccharomyces exiguus.				
ORGANISM	Saccharomyces exiguus.				
REFERENCE	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomyces.				
AUTHORS	1 (bases 1 to 1070)				
	Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G., Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S., deMontigny,J., Dujon,B., Durrens,P., Lepingle,A., Liorente,B., Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S., Saurin,W., Tekala,F., Toffano-Nioche,C., Wesolowski-Louvel,M., Winkler,P. and Weissenbach,J.				
TITLE	Genomic exploration of the hemiascomycetous yeasts: 1. A set of yeast species for molecular evolution studies				
JOURNAL	FEBS Lett. 487 (1), 3-12 (2000)				

REFERENCE	11152876	2 (bases 1 to 1070)	Bon, E., Neuveglise, C., Lepingle, A., Wincker, P., Artiguenave, F., Gaillardin, C. and Casaregola, S.
AUTHORS			
TITLE			Genomic exploration of the hemiascomycetous yeasts: 6.
JOURNAL			Saccharomyces exiguus
MEDLINE	20584716	FEB 5 Lett. 487 (1), 42-46 (2000)	
PUBMED	11152881		
REFERENCE		3 (bases 1 to 1070)	Genoscope.
AUTHORS			
TITLE			Direct Submission
JOURNAL			Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage, 2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail : segret@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT			This GSS is part of a random genomic sequencing program of thirteen yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii, Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila, Candida tropicalis and varrowia lipolytica. Genomic inserts of 3 to

5 kb were prepared and both extremities were sequenced. See keywords for description of this sequence and for the sequence of the other extremity of this insert.

```

FEATURES             source
    source            1..1070
                        /organism="Saccharomyces exiguus"
                        /strain="CBS 379"
                        /db_xref="taxon:34358"
                        /clone="AV0AA014A04"
                        /clone_lib="AV0AA"
                        /note="end : T3"
                        <11..>288
misc_feature         /note="similar to Saccharomyces cerevisiae ORF YLR330w [
                        CHS5 ; chitin synthesis protein ]
                        , 1 putative frameshift(s)
                        /evidence=not-experimental
BASE COUNT          356 a 218 c 245 g 221 t 30 others
ORIGIN
Query Match          14.9% ; Score 98.2 ; DB 17; Length 1070;
Best Local Similarity 42.1% ; Pred. No. 1.3e-09;
Matches 122; Conservative 25; Mismatches 143; Indels 0; Gaps 0;

QY 334 GTNCCNCCNWSNMGNTTYYTYSNGCNGCNGCNCNGCNCNCNCCNATHGCNCGN 393
      |||::: : |||::: |||::: |||::: |||::: |||::: |||::: |||:::
Db 592 GAACCTGCTACTRAGGAACMTGCTACTGAGGAACCTGCTACTGAGGAACCTGCTACTGAG 651
      |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
QY 394 GARCNGCNGCNGCNGCNCNCTYNACNCCNACNCCNGTNGCNGCNGARCCNGCNGCNGN 453
      |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
Db 652 GAAATGCTACTGAGGAACCTGWTACTGAGGAACCTGWTACTGAGGAACCTGCTACTGAG 711
      |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
QY 454 GCNCCNGTNGCNGCNGARCCNGCNGCNCNCTNGCNGCNGARCCNGCNGCNGAR 513
      |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
Db 712 GAACCACTGCCAGGAACCTGCTACTGAGGAACCTGWTACTGAGGAACCTGCTACTGAG 771
      |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
QY 514 GCNCCNGTNGCNGCNGARCCNGCNGCNGARCCNCCNCTNGCNGTNGCNGCNGCNGAR 573
      |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
Db 772 GAACCTGCTACTGACGAACCACTGCCGAGGAACCTGCTACTGAGGAACCTGCTACTGAM 831
      |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
QY 574 GARCNNSCNCCNGCNGARCCNGCNGCNCNACNCCNARCCNCCNCCNGARCC 623
      |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
Db 832 GAACCTCTATGTAGCAACCACTGCCGAGGAACCTGCTACTGAGGAACC 881
      |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::

RESULT 7
BF544627
LOCUS               BF544627                499 bp      mRNA      linear      EST 11-DEC-2000
DEFINITION          UI-R-BT0-qi-c-05-0-UI.r1 UI-R-BT0 Rattus norvegicus cDNA clone
                    UI-R-BT0-qi-c-05-0-UI 5', mRNA sequence.
ACCESSION            BF544627
VERSION              BF544627.1
KEYWORDS             EST.
SOURCE               Norway rat.
ORGANISM             Rattus norvegicus
                    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                    Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
                    Rattus.
REFERENCE            1 (bases 1 to 499)
AUTHORS              Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE               Normalization and subtraction: two approaches to facilitate gene
                    discovery
JOURNAL              Genome Res. 6 (9), 791-806 (1996)
MEDLINE              9704477
COMMENT             Contact: Soares, MB
                    Program for Rat Gene Discovery and Mapping
                    University of Iowa
                    451 Eckstein Medical
                    Tel: 319 335 8250
                    Fax: 319 335 9565
                    Email: msoares@blue.weeg.uiowa.edu
                    cDNA Library Preparation: M.B. Soares Lab Clone distribution:
                    clones will be available through Research Genetics (www.resgen.com)
                    This clone is also available through the I.M.A.G.E. Consortium at
```

ORGANISM

Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Chronic: 2006010801;

Submitted (12-Apr-2000)
This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at <http://www.genoscope.cns.fr/Tetraodon>.

[illegible]

Search completed: February 12, 2003, 12:36:30
Job time : 969.022 secs

